

Philips Technical Review

DEALING WITH TECHNICAL PROBLEMS
RELATING TO THE PRODUCTS, PROCESSES AND INVESTIGATIONS OF
THE PHILIPS INDUSTRIES

EDITED BY THE RESEARCH LABORATORY OF N.V. PHILIPS' GLOEILAMPENFABRIEKEN, EINDHOVEN, NETHERLANDS

THE "EXPRESSOR" SYSTEM FOR TRANSMISSION OF MUSIC

by R. VERMEULEN and W. K. WESTMIJZE.

534.86:621.395.665.1

In the transmission — either by radio or via a system of sound-recording — of the musical performance of a large orchestra the following difficulty is encountered. Whereas in the concert hall the differences in sound level may amount to about 80 db, the systems of sound transmission referred to, if equipped for a wide frequency band, cannot cope with any greater differences than, say, 50 db without introducing distortion or background noise arising in the reproduction. "Compression" is therefore commonly applied, which means that the loud passages are less amplified than the soft ones. Provided this compression takes place in the right way, for normal reproduction the result is reasonable, but dynamically it is of course far inferior to the original music. To improve matters in this respect systems of "expansion" have been devised with the object of compensating the compression in the reproduction. The present authors maintain that manual compression by a skilled expert is better than automatic compression. The "Expressor" system ensures that the degree of expansion is unambiguously determined by the degree of compression.

It is a well-known phenomenon that the reproduction of music is accompanied by more or less disturbing noises. Owing to the improvements that have gradually been introduced in the technique of reproduction the imperfections still remaining (among which is the presence of foreign noises) strike the eye — or rather the ear — more and more, with the result that the demands made of reproduction in this respect are becoming more and more severe. The case can be likened to the blacking out of a window: shutting off the large openings makes small cracks all the more noticeable. An example may illustrate what is meant.

One of the most outstanding recent improvements in the field of the transmission of sound ¹⁾ is stereophonic reproduction, whereby, according to the system developed by Philips ²⁾, the sound from, say, two loudspeakers, each connected to a

microphone or a pick-up, is caused to merge into one single "sound picture". The listener has the impression that this sound picture is somewhere between the two loudspeakers. This does not apply however to the background noise (insofar as this comes from the transmitting apparatus), since the noise contributions from the two loudspeakers are independent of each other and consequently the ear localizes them in the direction of each of the loudspeakers. Thus the noise is heard as coming from directions different from that of the music, so that it becomes easier to pay no attention to those noises. It might therefore be expected that with stereophonic reproduction the background noise would be less troublesome than in the case of ordinary reproduction. In conformity with what has been said in the previous paragraph, however, the reverse is often the case: the illusion of hearing original music, to which the stereophonic effect so strongly contributes, is repeatedly broken on account of the background noise reminding one that the music heard is only a reproduction.

As a matter of fact much depends upon the ability and the willingness of the listener to concentrate upon the music and thus not to listen to the

¹⁾ By transmission is to be understood here not only transmission from one place to another (via telephone lines or by radio) but also the production of a sound record, as for instance a gramophone record, a Philips-Miller tape, etc. for subsequent playing.

²⁾ K. de Boer, Stereophonic sound reproduction, Philips Techn. Rev. 5, 107-114, 1940; The formation of stereophonic images, Philips Techn. Rev. 3, 51-56, 1946.

background noise. There are not many listeners prepared to make any effort in this respect, as is evidenced by the position usually occupied by the tone control of radio-sets in the homes of radio listeners. These obviously prefer the loss of the high-pitched notes in the music to the disturbing effect of the background noise. The reason for this lies mainly in the common aversion to the hissing that forms one of the principal components of the background noise. Naturally, when the tone control is turned back the high notes are for the greater part lost. Although technical experts disagree on the question if and in how far the public can appreciate the faithful reproduction of the entire audible frequency range, the suppression of noise at the cost of the high notes of the music is generally considered to be an unacceptable makeshift.

The noise problem occurs in many fields of communication technique, but not everywhere to the same extent. In the case of telegraphy and telephony for instance it is a matter of transmitting the signals economically in such a way that they can be recognized with sufficient certainty in spite of interference. For the transmission of music, however, much more is demanded: the interferences must be at least unobtrusive, if not imperceptible. Lowering the level of the interferences below the auditory threshold is so difficult, in view of the enormous sensitivity of the ear, that it is to be doubted whether any success will ever be attained in that direction. Fortunately, however, there is another property of the ear that tends to act as a counterbalance, and that is the difficulty of distinguishing a weak sound in the presence of loud ones. Thanks to this property of the ear it suffices to attenuate the noise to such a level that it becomes masked, either by the music itself or by other sounds always present both in a concert hall (or studio) and in the auditorium where the music may be reproduced. These noises, apart from those penetrating into the hall from the outside, come from the musicians and the audience itself. This "hall noise", as we may call it, should, objectively speaking, be classified under the disturbing sounds, but, as a "natural" noise it is more acceptable to the listener than the noise of the reproduction system which it helps to mask. In fact it even contributes towards creating the desired atmosphere in the hall where the music is being reproduced.

Level ranges

What is the range of intensity levels possible between the hall noise and the loudest music

on the one hand and the range between the noise of the transmission apparatus itself and the maximum signal it can handle on the other hand?

The highest peaks occurring in the sound produced by a large orchestra reach close to the "feeling level" of the human ear. This is borne out by measurements by American investigators³⁾, who found that an orchestra of 75 musicians may produce in the peaks a power of about 70 W, which in the average concert hall corresponds to a level of about 110 db above the usual zero level of 10^{-12} W/m².

Measurements of the intensity of audience noise taken in cinemas⁴⁾ show that the sounds penetrating from the outside into an empty hall reach an intensity of 25 db above the zero level, while in halls where an audience is present the level of the noise varies between 38 and 44 db (average 42 db), dropping to 32 db during particularly thrilling parts of the film. The latter figure will not be far off the level to be reckoned with in a concert hall. Passages of music below this level will be drowned in the hall noise. Thus the difference between the highest and the lowest sound level in a concert hall — the "range" — amounts to about 80 db.

In the transmission of sound one has to take into account, on the one hand, the danger of non-linear distortion which threatens to arise at the high peaks in the intensity of the sound, and on the other hand noises inherent in the transmitting system. In the case of the Philips-Miller tape⁵⁾, to which the following considerations will be mainly confined, a difference of at most 50 db can be reckoned with between the noise level and the level at which this non-linear distortion becomes noticeable. (The former also depends on the width of the frequency spectrum, which is taken here as being 8000 c/s.)

As a rule the intensity range afforded by the transmitting system is not wide enough to cope with the intensity range of the original music; in our case it is 30 db too short.

Levels determining the intensity range in transmitting systems

Before discussing the measures that have to be taken to meet this discrepancy we have to go further into the factors determining the intensity range of a system with optical scanning, like that employed for sound films. Let us first consider the lower limit, which is determined by the background noise.

- ³⁾ L. J. Sivian, H. K. Dunn and S. D. White, Absolute amplitudes and spectra of certain musical instruments and orchestras, *J. Acoust. Soc. Amer.* **2**, 330-371, 1931.
- ⁴⁾ W. A. Mueller, *J. Soc. Mot. Pict. Engrs.* **35**, 48-53, 1940.
- ⁵⁾ R. Vermeulen, The Philips-Miller system of sound recording, *Philips Techn. Rev.* **1**, 107-114, 1936.

Between the original and the reproduced sound there are the following links: microphone - amplifier - sound recorder - sound tape - light ray - photocell - amplifier - loudspeaker, each of which contributes its share in the total noise. These sources of sound are incoherent, so that the powers of each noise have to be added and not the amplitudes. The contribution of a source with an amplitude say half of that of another source is thus only 1 db. Therefore, the strongest source of noise practically determines the total noise level.

When modern microphones are used in a suitable circuit and the amplifiers are of a carefully designed construction these elements are certainly not the main source of the background noise, and neither are the loudspeakers.

The photocell calls for rather more attention. The noise of the photocell results from the fact that the electrical signal is the integrated effect of many electrical particles (electrons) and as such this noise is proportional to the square root of the number of particles passing through the cell per second, and is thus proportional to the square root of the average value of the photocell current. This average current is limited at the lower end by the fact that it must always be at least equal to the peaks with which the current is modulated by the sound track of the film. There are in fact systems for counteracting noise (commonly termed "noiseless") whereby the average photocell current is varied so as not to be greater at any time than what is required for the modulation, the photocell noise thus being minimized. Any further improvement could only be obtained if it were possible to increase the illumination of the track. If this illumination could be doubled, for instance by using a lamp with greater brightness or by employing a better optical system, the signal amplitude at the output of the photocell would likewise be doubled but the noise would only increase by a factor $\sqrt{2}$, so that the intensity range between the maximum signal and noise would become greater by a factor $\sqrt{2}$, i.e. 3 db. With the apparatus commonly employed however any improvement of this nature is hardly to be expected, nor is it urgently needed.

The light, too, has a "granular" structure (photons), so that the beam of light striking the photocell forms in itself a source of noise. Since, however, on an average about 200 photons are required to release one electron from the usual photo-cathode, the structure of the electron current flowing through the photocell is much coarser than that of the beam of light, and the noise contribution of the latter may be ignored.

Finally there is the noise of the sound track. This is stronger than that of any of the sources summed up in the foregoing, even in the case of a freshly cut Philips-Miller tape, although, in comparison with other systems for optical scanning, its inherent noise is very low. This low noise level of the Philips-Miller tape results from the fact that the cutter entirely removes the opaque coating from the sound track (see fig. 1a), while outside the track the amount of light passing through is quite negligible (in contrast with photographic films, where the grained structure of the emulsion gives rise to noise; see for instance figs. 8 and 12 in the article quoted in footnote 5). The main source of the low tape noise lies in the narrow wedge-shaped edges of the track (*A* in fig. 1b), where some fluctuation is possible in the density. To this are to be added the irregularities in the cutting process itself. When it is borne in mind that the edges of the track are formed by the cutter tearing away the material it is really surprising that it is possible to get the excellent definition of the edges of the track that is needed for this purpose.

More serious is the noise that is to be ascribed to quite commonplace causes but which often exceeds all other contributions, namely the noise due to particles of dust and scratches or indentations on the surface of the film and to impurities in the material of the tape. Specks of dust and impurities intercept, and scratches scatter light, all this resulting in a kind of noise that could best be called "sputtering".

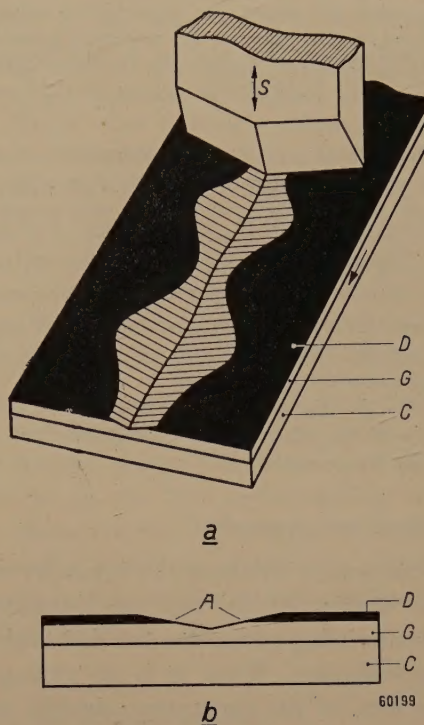


Fig. 1. a) The recording of sound on a Philips-Miller tape. This tape consists of a strip of celluloid *C*, on which a layer of transparent gelatine *G* and a thin opaque coating *D* are applied. The tape moves along underneath the cutter *S* in the direction indicated by the arrow, the cutter moving up and down, thereby cutting a sound track in the opaque coating. b) Enlarged cross section of a Philips-Miller tape; *A* = edges of the sound track; the other letters have the same meaning as in (a).

Enlarging the intensity range by raising the upper limit is rendered difficult by a number of factors, some of which are of a fundamental nature while others are of a more practical nature. The Philips-Miller recorder could be made to yield larger amplitudes without distortion by giving it a lower resonant frequency. (To avoid a rise in noise owing to the track then necessarily being wider, the above-mentioned "noiseless" system could be applied.) This greater amplitude, however, can only be obtained at the cost of the high-frequency response⁶).

We would mention here another method, likewise not without its drawbacks, which would in theory enable the intensity range to be enlarged. When the sound spectrum is imagined as being divided into a number of frequency

⁶ With the Philips-Miller recorder developed for stereophonic recording (two tracks on one tape) the opposite course has been followed: in order to get a flatter frequency characteristic a narrower track is used. The two stereophonic tracks together take no more room than the former single track.

bands it is found that the large amplitudes do not usually occur in all bands. Given this distribution (which is known for most musical instruments and for speech in different languages), the recording installation can be given such a frequency characteristic that the bands in which normally only small peaks occur are amplified more than the others, so as to make it equally probable for the peaks in each band to reach the limits of the sound track. To obtain a reproduction with a flat overall characteristic, this difference in amplification must be compensated by giving the reproducing amplifiers the inverted frequency response of the recording apparatus. This means that the frequency bands with the small signals which were given extra emphasis in recording are now correspondingly attenuated.

The result is then a similar attenuation of the noise contribution of those bands. Obviously this measure can only be successful if in certain bands there is only a weak "signal" (weak components of the music or speech) — in relation to the maximum amplitudes — and much noise. In favourable circumstances (orchestral music) the intensity range can in this way be widened by 10 db⁷⁾; but for speech, at least in strongly sibilant languages, this method does not hold. Another drawback is that owing to its special frequency characteristic the reproducing apparatus is not suitable for those forms of recordings where this method has not been applied, and vice versa.

Compression and expansion

The discrepancy between the intensity range in orchestral music (80 db) and the intensity range of the transmitting apparatus (50 db) leads to the following difficulty. When it is so arranged that the noise level of the apparatus coincides with the level of the hall noise, i.e. 30 db, (respectively C_1 and A in fig. 2), all peaks in the music ascending

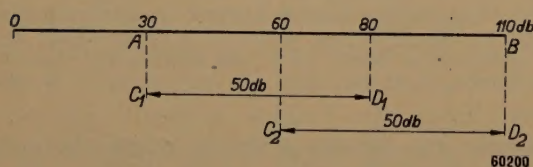


Fig. 2. A = level of the hall noise (30 db), B = level of the highest peaks in orchestral music (110 db) above the zero level 0 (10^{-12} W/m²). If the 50 db range of an installation for sound transmission is laid along C_1D_1 then distortion arises at all peaks higher than 80 db. If the range is laid along C_2D_2 then the noise level (60 db) will be high.

higher than 80 db above the zero level are overmodulated. When the limit for the maximum permissible level of the apparatus is made to coincide with the highest sound level (respectively D_2 and B in fig. 2) no distortion will occur, but the noise level (C_2) becomes so high as to drown all music not reaching 60 db above zero level.

The remedy usually followed to overcome this difficulty consists in reducing the intensity range of the music to 50 db by what is known as "compression" — for instance by means of a volume control operated either by hand or automatically — thus attenuating the loud passages compared with the soft ones⁸⁾. These reduced differences in level are familiar to everyone who listens to radio and gramophone music.

Musicians quite rightly object to their work being treated, or rather ill-treated, in this way. Although "compressed" music may be acceptable provided the compression is carried out by a skilled expert, different means have been sought to restore the original contrasts in the reproduction, so long as there is no alternative to compression. These means are all based upon the application of expansion, restoring the original relations in the music in some way or other by giving more emphasis to the high levels, thereby compensating the compression. The expander used for this purpose may be operated automatically, for instance by causing it to respond to the amplitude of the signal fed to it⁹⁾.

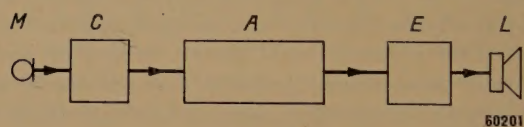


Fig. 3. Diagrammatic representation of an installation for sound transmission. M microphone, C compressor, A complex of amplifiers and any recording and reproduction apparatus that may be used, E expander, L loudspeaker.

The chain between microphone and loudspeaker is then as represented in fig. 3, where A is a complex of amplifiers and possibly a recording and scanning device. An important point, as we shall presently see, is the fact that the main sources of noise are contained in A , that is between the compressor (C) and the expander (E).

Shape of the compression characteristic

For the moment we shall defer a consideration of the question whether compression can be carried out better by hand or automatically, and assume for the time being that it is done automatically, according to a fixed relation between the degree of attenuation C , introduced by the compressor, and the intensity I_1 of the original music. It will be taken

⁷⁾ J. C. Steinberg, The stereophonic sound-film system; pre- and post-equalization of Compandor systems, J. Soc. Mot. Pict. Engrs. 37, 366-379, 1941, in particular p. 374.

⁸⁾ See, e.g., R. Vermeulen, The relationship between fortissimo and pianissimo, Philips Techn. Rev. 2, 266-269, 1937.

⁹⁾ See, e.g. V. Cohen Henriquez, Compression and expansion in sound transmission, Philips Techn. Rev. 3, 204-210, 1938.

for granted that when I_1 rises from 30 to 110 db the attenuation must increase from 0 to 30 db¹⁰). Thus in *fig. 4a* the initial point *A* and the final point *B* of the compression characteristic are fixed.

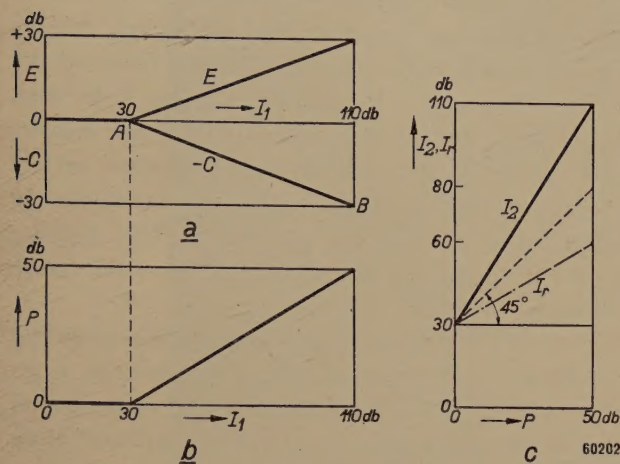


Fig. 4. *a*) The compression C and the expansion E plotted as functions of the intensity I_1 of the sound in the concert hall. It is assumed that the compression increases linearly according to AB when I_1 is increased from 30 to 110 db. To restore the original dynamic contrast it is necessary that $E-C=0$. *b*) The amplitude P of a sound record made with the compression characteristic according to (*a*) plotted as a function of I_1 . *c*) Plotted as functions of P : the intensity I_2 of the reproduction (dotted line for the case where expansion is not applied, full line for the case where expansion is applied), and the intensity I_r (dot-dash line) of the noise level (with expansion).

The shape of the characteristic between *A* and *B* is to a certain extent arbitrary. The simplest characteristic is the straight line AB (*fig. 4a*). The amplitude P of the recorded signal¹¹) then assumes a shape, as a function of I_1 , as represented in *fig. 4b*, whereby the range of 80 db in I_1 is reduced to 50 db in P . If no expansion is applied in the reproduction then the intensity I_2 of the reproduced music, as a function of P , follows the dotted line in *fig. 4c*, with the same range (50 db) as P . By applying expansion the level of the reproduced sound I_2 can be given a range equal to that of I_1 (full line in *fig. 4c*); the expansion characteristic $E=f(I_1)$, *fig. 4a*, must then be the inverted image of the compression characteristic ($E-C=0$ for every value of I_1).

As already pointed out, the main sources of noise are located between the compressor and the expander (in *A*, *fig. 3*). The noise level I_r of the reproduction therefore follows the movements of

the expander (see the dot-dash line in *fig. 4c*) and as these have to follow the movements of the compressor (in order to maintain the relation between the two characteristics, $E-C=0$) the noise in the reproduction therefore varies with the intensity of the music. This is most unpleasant, for, in spite of the masking effect of the correspondingly varying intensity of the music, a varying noise attracts attention much more than a constant noise, from which one can detach one's mind even if it is rather strong. Hence the question arises whether the compression characteristic represented in *fig. 4a* is the most favourable, since according to that characteristic compression is already applied (and noise is thereby emphasized and varied) as soon as I_1 rises above the level of 30 db, where there is no danger of overloading whatever. It is therefore obvious that it would be more favourable if the compression were made to start at a higher level of I_1 , in fact at such a level that the noise is already masked by the music, for instance at 70 db, as indicated in *fig. 5a*. The corresponding variations of the characteristics $P=f(I_1)$ and $I_2=f(P)$ (*figs 5b* and *c*) need no explanation. From *fig. 5c* it is seen that the noise level is now constant over a large range of P .

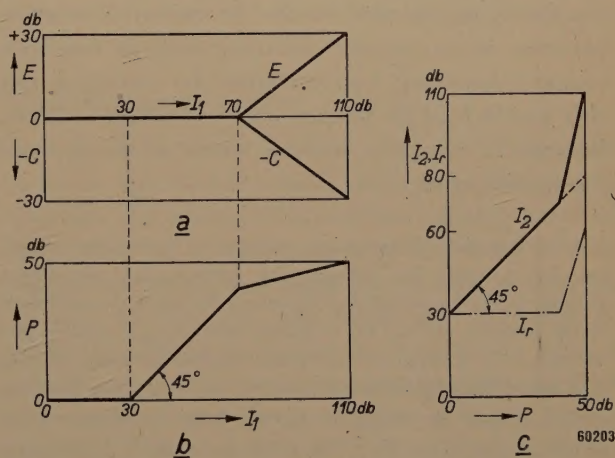


Fig. 5. The same as *fig. 4* but with a compression characteristic starting at $I_1=70$ db.

Is it possible to go still farther in this direction? Strictly speaking, there is no need for compression so long as there is no danger of overmodulation, i.e. before P reaches 50 db, corresponding to $I_1=80$ db above zero level. The characteristics then assume the shape given in *figs 6a, b* and *c*. A difficulty then arises, however, in that, in the vertical part of the characteristic in *fig. 6c*, I_2 is then no longer a single-valued function of P . In other words,

¹⁰) This figure of 30 db agrees with what H. Fletcher gives for variable-area sound-film records (*J. Soc. Mot. Pict. Engrs.* 37, 331-352, 1941, in particular p. 338). Some of the ideas developed here have been taken from this article by Fletcher.

¹¹) Where no record is made, as in the simultaneous reproduction of music elsewhere, P may be taken as representing the signal at the input of the expander.

contrary to the cases of figs 4 and 5, the expander cannot of itself deduce from the amplitude of the input signal P what the expansion has to be. A solution for this problem will be indicated presently.

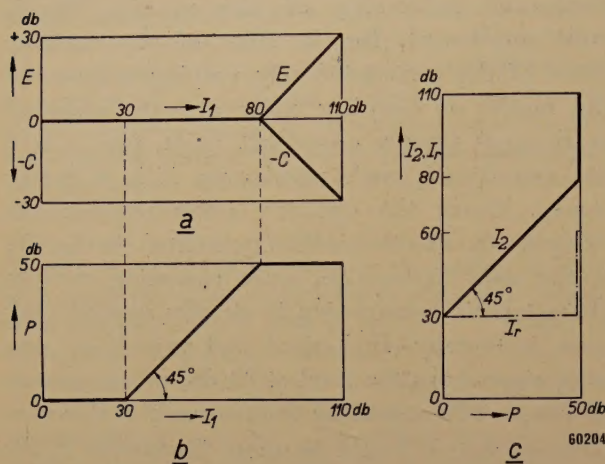


Fig. 6. The same as fig. 4 but with compression starting at $I_1 = 80$ db, thus at the level at which, without compression, distortion would arise. At higher values of I_1 there is no longer any unambiguous relation between I_2 and I_r on the one hand and I_1 on the other hand.

Automatic or hand-regulated compression?

It is impossible for the relation between the compression and the level of the music to be so simple as represented in the foregoing. A relation between momentary values, such as has been roughly assumed, can certainly be obtained but this would lead to strange and undesirable effects, because it would in essence mean a strong non-linear distortion.

A compression characteristic relating to momentary values can be realized, for instance, by feeding the microphone voltage to the control grid of a pentode having a characteristic of the shape $i_a = c_1 v_g^n$ (i_a = anode current, v_g = control-grid voltage, c_1 a constant, n a number between 0 and 1). The non-linear distortion then obviously occurring can be eliminated, at least in theory, by using for the expansion a valve having a characteristic of the shape $i_a = c_2 v_g^{1/n}$ (where c_2 is again constant). It is then essential, however, that the factors n in the two exponents have exactly the same value. Another requirement, just as difficult to meet, is that no phase shifts may occur between the components of the signal on its way from the compressor to the expander. In practice such a system would result in strong distortion being heard not only in the highest levels (as would be the case without any compression) but almost continuously.

To avoid this distortion the variations in the attenuator can be made to take place so gradually as to allow the system always to be quasi-stationary with respect to the sound waves, but the compressor would then be unable to fulfil its task properly. This

task is on the one hand to avoid overloading and on the other hand to avoid too high a noise level. If, however, the compression has to lag somewhat behind the variations in intensity of the music, then upon the sudden starting of a fortissimo passage there is bound to be a clipping, and an abrupt pause of the music will be followed by noise dying away. In the latter case, it is true, owing to reverberation the sound in the concert hall will die away gradually and thus more or less mask the noise, but, nevertheless, variations in noise level are likely to be noticed. The clipping at the sudden rise of the level is even more annoying.

Various methods have been suggested for avoiding this distortion arising when an automatic compressor is used. In the first place the compression can be made to start at a slightly lower level than where overmodulation occurs, for instance as indicated in fig. 5a, thus leaving some margin for the first peaks of a sudden fortissimo; this margin, which would have to be about 6 db, could only be obtained, however, at the cost of the intensity range of the recording, because normally it cannot be used.

Other suggestions are based upon the idea of bringing the automatic compressor into action just before the first peak of the fortissimo reaches the recording apparatus. This idea can be realized by placing between the orchestra and the microphone an auxiliary microphone for picking up the sound earlier and operating the compressor before the same sound has reached the main microphone. The objection against this solution, however, is that it sets intolerable limitations to the choice of the microphone positions. Moreover, the interval of time that can be obtained in this way (or alternatively, for instance, with a delaying network between the compressor and the recording apparatus) is at most of the order of a few milliseconds. If clipping of the peaks is to be effectively avoided the compressor would have to act so rapidly that no quasi-stationary state could possibly be reached and, therefore, distortion at low levels would replace the clipping.

A third method, frequently applied with automatic compressors, consists in causing the compressor to respond as quickly as is possible (limited for other reasons) to an increase of the sound level but to lag behind a decrease in sound level. The first peaks of a fortissimo passage then still cause clipping, but this is only of short duration, owing to the now almost instantaneous response of the compressor. Any delay in the

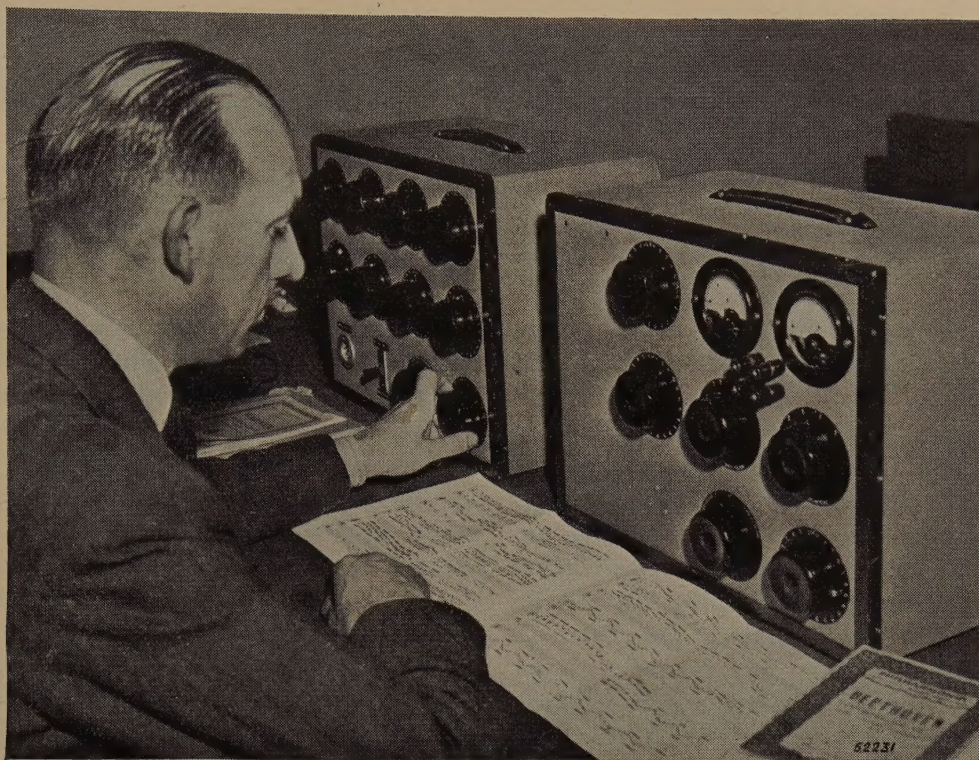


Fig. 7. The "music pilot" at work. He regulates the compression guided by the music score in front of him. By means of the other controls seen in the photograph the signals from the microphones set up in the hall can be mixed in the right proportions.

compensation by the expander also lasts too short to be noticeable. The time constant of the delay at a sudden drop of the level has to be so chosen that the "tailing on" of noise is the least troublesome. In this way a compromise can be reached where an untrained ear scarcely notices the momentary distortion or the varying noise.

In our opinion, however, such an automatic system avoids the issue and it is by far preferable for the compressor to be operated by hand, but then by a capable hand! No mechanism has the power of prediction that is indispensable if it is to take action in good time and in the right way before a peak arises. Only one trained in music has the foreknowledge — especially when he has the score in front of him — which, together with a good dose of routine, is necessary to be able to reduce the amplification gradually and in good time before a fortissimo passage starts, and to increase the amplification in the right way upon the termination of a loud passage. To reach the ideal solution, sense and an ear for music are indispensable ¹²⁾.

In the majority of cases compression is in fact

regulated by hand (see fig. 7). When listening to radio, gramophone or film music anyone can judge how satisfactory the result can be, provided the operator is skilled in his job.

The "Expressor" system

Once it has been decided that the compressor is to be operated by hand there is no longer any question of a certain fixed relation existing between the degree of compression and the level of the original music as was supposed to exist in figs. 4a, 5a and 6a; the personal views and the skill of whoever may be operating the compressor play a part. To put it in other words, the amplitude P of the sound record is no longer a single-valued function of the sound level I_1 . If expansion is to be applied then a difficulty arises similar to that encountered in the case of fig. 6c, where I_2 is no longer everywhere a single-valued function of P — likewise arising from the fact that I_1 is not everywhere a single-valued function of P (see fig. 6b). This difficulty lay in the fact that as soon as these fixed relations are broken the expander can no longer deduce from the signal P what level I_1 actually had (and thus what the level I_2 should be). A solution of this difficulty is the following: the position of the compressor at any moment can be communicated to the expander via a separate

¹²⁾ The experiments which will be mentioned presently were made possible by the devoted and capable cooperation of Mr. M. J. C. van der Meulen (Electro-Acoustical Dept., Eindhoven).

transmission channel and the expander can be made to compensate the compression continuously. In more technical terms, through the intermediary of a pilot signal the compressor has to control the expander in such a way that $E-C$ is always equal to zero.

We have given this system the name of "Expressor", which is made up from parts of the words expander and compressor and at the same time signifies the greater possibilities of expression which can be obtained by this system in the reproduced music.

When listening to the monitoring loudspeaker one will not notice in the music any influence whatever of the adjustment of the control of the "Expressor", apart from the fact that turning the knob in one direction results in an increasing distortion, while the noise level steadily increases in the other direction. The readings of the meters on the control panel indicate how the level of the output signal of the compressor is varied and how it should be adjusted.

There are two points of view from which one can consider the operation of the "Expressor". The simplest way is to disregard the expansion and to pilot the signal in the usual way between noise and distortion. In the case of reproduction without expansion the result will have the same dynamic character as ordinary radio and gramophone music; with expansion the original dynamic nuances are fully restored. Operated in this way, however, the inherent possibilities of the "Expressor" are not utilized to the full. This is only the case when the noise level is continuously kept as low as possible, that is to say, when the sound track is continuously modulated close up to the limit of distortion. This imposes a heavier task upon the "music pilot" operating the compressor, for he has to be on the alert all the time and can scarcely do without the music score in front of him, or even a rehearsal, if he is to take action in good time for every crescendo. The reward for such painstaking attention is a reproduction which only exceptionally reminds one of the technical element involved. Particularly striking is the absence of noise during the "rests" and the pianissimo passages. Only in the fortissimo passages, if one concentrates on it sufficiently, can any foreign noise be detected. This is just the opposite of what one has become accustomed to in the reproduction of music, and in our opinion it means an appreciable improvement.

For safety's sake an automatic auxiliary compressor might be added to come into action only when there is a risk of overloading. This seems to be a wise precaution, but we have

not adopted this measure because it was feared that the "music pilot", relying on this automatic compressor, would not be induced to concentrate to the full on his exacting work. It has been found that if the pilot takes his work seriously there is no need for any such precaution.

Technical execution

It now remains to discuss the method chosen for controlling the expander by the position of the compressor.

The music pilot operates a potentiometer controlling the degree of compression. The expander consists of another potentiometer which has to respond to every change in the compression to a like degree in the opposite sense. (Of course for stereophonic transmission two compression and two expansion potentiometers are needed.) The problem of effecting this coupling (in space and time!) between the compression and the expansion potentiometers is possible of solution in various ways. Here only a brief description will be given of an impulse system which was applied for the experiments carried out in the Concert Hall in Amsterdam in 1947¹³).

With stereophonic transmission the compression

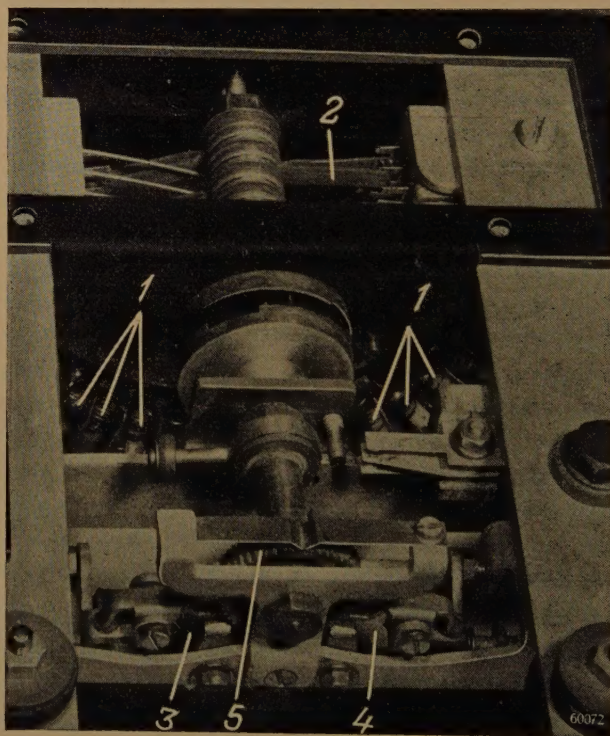


Fig. 8. Mechanism of the expander. At 1 are to be seen some of the resistors which form a potentiometer and are connected to a series of contacts traversed by a contact brush 2. This is turned in one direction or the other by the action of the armature of the electromagnet 3 or 4 causing the ratchet wheel 5 to turn one or more steps.

¹³) R. Vermeulen, Duplication of concerts, Philips Techn. Rev. 10, 169-177, 1948 (No. 6).



Fig. 9. Illustration (twice the actual size) of a strip of Philips-Miller tape with two stereophonic sound tracks. Underneath these tracks are marks (pilot track) made as the result of the compression potentiometer having been moved two steps in a certain direction. When scanned during the play-back these marks cause the expansion potentiometer to be rotated the same number of steps in the opposite direction. Above the sound tracks similar marks can be made for turning the expander potentiometer in the other direction.

has to be adjusted in the two channels in exactly the same way by means of one single knob (*G* in fig. 1 in the article just quoted), so as to avoid any displacement of the stereophonic sound picture. Consequently the potentiometers used have to be of a very high quality. The most reliable potentiometers are variable in steps with wire-wound resistors. If the steps are no more than 1 to 2 db the ear

does not perceive the discontinuities in the adjustment.

The compression potentiometers are fitted with an auxiliary contact which at every step produces an impulse that is transmitted via the pilot channel to the expandors, where it activates an electromagnetically operated ratchet gear (fig. 8) which adjusts the expansion potentiometers in the desired

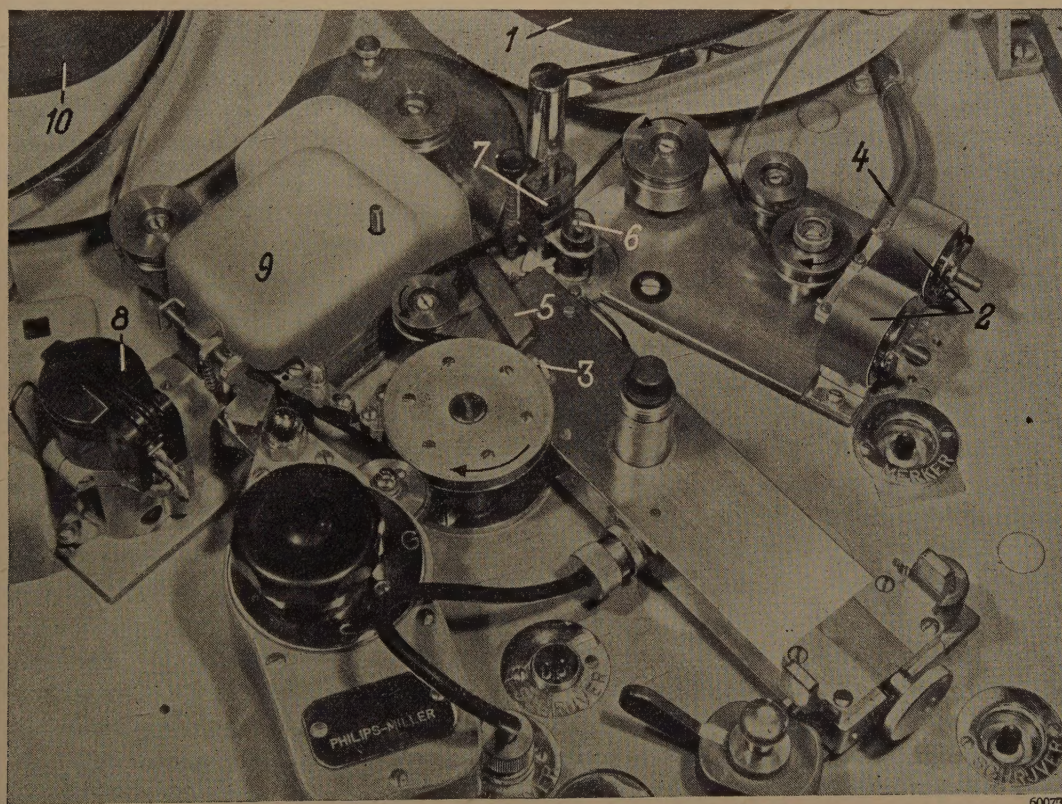


Fig. 10. The principal parts of a Philips-Miller machine with "Expressor" system. This machine is suitable both for recording and for play-back. For recording, 1 is a roll of blank Philips-Miller tape, passing along in front of the markers 2, which cut the pilot track indicating how many steps and in which direction the compressor has been moved. It then passes by the cutters, only one of which (3) is to be seen here; the other, which should be on top of 3, has been removed so as not to obscure the first one. At 4 and at 5 are suction tubes for the swarf (the material cut out of the tape). The tape is rewound at 10. If so desired, the music can be played back while the recording is still proceeding. The tape is then passed — as shown in the picture — along the optical scanning devices, one for the pilot tracks and the other for the sound tracks. The exciter-lamp of the former scanning device is to be seen at 6; the photocells (selenium cells) are at 7. The exciter-lamp for the sound-track is at 8 and the two photocells are at 9. For the expander to be brought into action at exactly the right moments the length of the tape between 2 and 3 must be equal to that between 6-7 and 8-9. When a previously made tape record is to be played then of course the tape has only to pass along 6-7 and 8-9.

direction. Naturally the impulses corresponding to one direction of rotation of the compression knob have to be distinguished, or kept separate, from those for the other direction of rotation.

This can be done, in the case of simultaneous reproduction, for instance by modulating each of the two series of impulses on a carrier of its own and transmitting these carriers to the expander via a telephone line or by radio. For recording on Philips-Miller tape the space at the sides of the sound-track offer room for two "pilot tracks", each of them containing a series of impulses (*fig. 9*). In *fig. 10* the cutting units ("markers") for these two pilot tracks are to be seen at 2, and their optical scanners at 6-7.

By choosing impulses for the pilot signal its exact amplitude becomes of no consequence and spurious voltages in the pilot channel will not readily give rise to errors. The only requirement that has to be fulfilled is that the contrast between the impulses and spurious voltages is sufficient and that successive impulses do not overlap.

When the steps of the compression and expansion potentiometers are logarithmic it is not necessary for these potentiometers to be in corresponding positions, but this is nevertheless desirable because

otherwise the range covered by the two together is reduced. To be able to bring them easily into corresponding positions the expander potentiometers are provided with a stop in the extreme positions; all one need do, therefore, is to turn the compressor knob from one extreme position to the other just before the transmission of the music begins, to obtain the desired correspondence in position of the potentiometers.

Summary. In systems for sound transmission with a wide frequency band the difference between the level at which overloading starts and the level of background noise is less than the difference that may occur in orchestral music between the loudest passages and the noise in the concert hall, so that one is faced with the necessity of applying compression in order to avoid both overloading and high background noise. In the reproduction of the music the original dynamic differences can be regained by means of expansion. — It is claimed that adjustment of the compression by a capable hand has great advantages over automatic control. With non-automatically adjusted compression however there is no unambiguous relation between the intensity of the original music and that of the input signal of the expander, and consequently special measures have to be taken to ensure that the expander always exactly compensates the compression. In the "Expresor" system a pilot signal causes the potentiometer of the expander to follow continuously the movements of the compressor. This pilot signal is transmitted via a separate channel and may consist, for instance, of impulses. If Philips-Miller tape is used as a sound record the pilot signal can be registered on the same tape in two pilot tracks (one for each direction of rotation of the compressor).

PHANTOM TESTS WITH X-RAYS

by G. C. E. BURGER.

778.33:621.386.1:
616-073.75:771.534.5

The medical practitioner wishes a lung radiograph to be capable of revealing details which are hardly perceptible. If he does not see a certain suspected detail does this then mean that that detail is non-existent, or is it that the X-ray picture is unable to reveal it? For a study of the general problem of perceptibility of details in X-ray technique good use can be made of specially constructed phantoms.

In the formation of an X-ray image there are a large number of factors playing a part. Besides the absorption and the scattering of the X-rays in the object there are the properties of the photographic film or, in fluoroscopy, those of the fluorescent screen and, furthermore, also the properties of the optical system (camera) if such is used.

The quality of the image can be judged according to its two main characteristics, definition and contrast of details. Contrast is defined by the ratio of the brightnesses of the detail and the background. A small detail is more readily observed when it stands out clearly against the background, that is when the contrast is great, than when such is not the case. Definition of details depends upon the X-ray tube used (dimensions of the focus) and, to a large extent, upon the fineness of grain of the photographic film or the fluorescent screen. Contrast is mainly determined by the physical properties of the object itself, by the voltage on the X-ray tube and by the gradation of the film.

In this paper attention will be specially directed to medical radiology and in particular to its application in the X-raying of the lungs. In practice different methods are followed for radiographing the lungs, the most important of which are:

- 1) contact radiography, whereby the X-rays, after passing through the object, impinge directly on the film, which may or may not be provided with one or more intensifying screens;
- 2) fluoroscopy, whereby the X-rays impinge on a fluorescent screen and the image thus formed is viewed visually;
- 3) fluorography; a camera is set up behind the fluorescent screen and the image is photographically recorded on a film on a reduced scale.

With all three methods it is possible to obtain a certain amount of magnification by placing the object at some distance in front of the screen or film.

It is difficult to make a quantitative comparison

between these methods and between the various ways in which one particular method may be carried out, owing to the previously mentioned large number of factors playing a part in the formation of the X-ray image. The introduction of the phantom in X-ray technique is a great improvement in objective testing. In this article it will be explained how a comparison between the various methods can be arranged on a quantitative basis with the aid of phantoms.

Description of a phantom. Contrast-detail diagram

In its simplest form a phantom consists of a plate of some kind of material the thickness of which varies in a known manner. By radiographing this plate the value of the radiographic method used can to a certain extent be determined from the differences in brightness of the image. Tests with such a phantom have been carried out by Bronkhorst¹⁾ and Luft²⁾, among others.

It is not possible, however, to determine with such a phantom any relation between the visibility on the one hand and the contrast and definition of detail on the other, because the phantom does not possess any small details. Phantoms having small details have already been commented in this journal³⁾⁴⁾. Here a description will be given of the phantom that has been used for the investigation dealt with in this paper and which in principle resembles the phantom that was used for the article quoted in footnote⁴⁾.

The phantom consists of a number of plates of

1) W. Bronkhorst, Kontrast und Schärfe im Röntgenbild, published by Thieme, Leipzig, 1927.

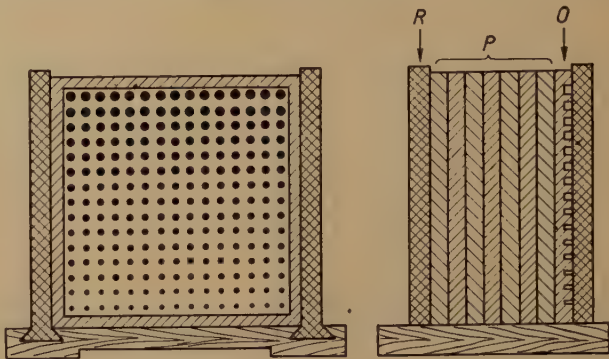
2) F. Luft, Experimentelle Beiträge zur Detailerkennbarkeit und Detaildarstellbarkeit bei verschiedener Aufnahmetechnik, Fortschr. Röntgenstrahlen 51, 412-417, 1935.

3) B. van Dijk, Several problems of X-ray fluoroscopy, Philips Techn. Rev. 4, 114-117, 1939.

4) G. C. E. Burger, B. Combée and J. H. van der Tuuk, X-ray fluoroscopy with enlarged image, Philips Techn. Rev. 8, 321-329, 1946.

“Philite”, one of which serves as object plate (fig. 1) and has 225 holes drilled in it. The holes in each horizontal row are of the same diameter but vary in depth between 8 mm and 1 mm in 15 steps of 0.5 mm; those in each vertical column are equal in depth but vary in diameter, likewise between 8 mm and 1 mm in steps of 0.5 mm. These holes serve as details, the visibility of which is investigated in the X-ray image. The other “Philite” plates can be inserted in or taken out of the phantom as desired for imitating the thickness of the thorax, the aim being to collect data about the methods applied in practice for pulmonary diagnostics. A constructional diagram of the phantom is given in fig. 2. To make it as near as possible comparable to an actual radiograph of the lungs the phantom can be placed in an apparatus like that depicted in fig. 3, with which the movements resulting from the heart beat of the patient can be imitated; the phantom is moved by means of a pendulum and the rate of movement can be adjusted by varying the maximum deflection. A small lead plate that can be attached to the phantom makes it possible to measure exactly the motional unsharpness and thus the exposure time.

“Philite” has been chosen as the material for the phantom because it most closely resembles the lung tissue as regards its properties determining



60261

Fig. 2. Constructional diagram of the phantom. It comprises the object plate O and a number of plates P without details; all plates are of “Philite”. The frame R can be moved to take a varying number of plates.

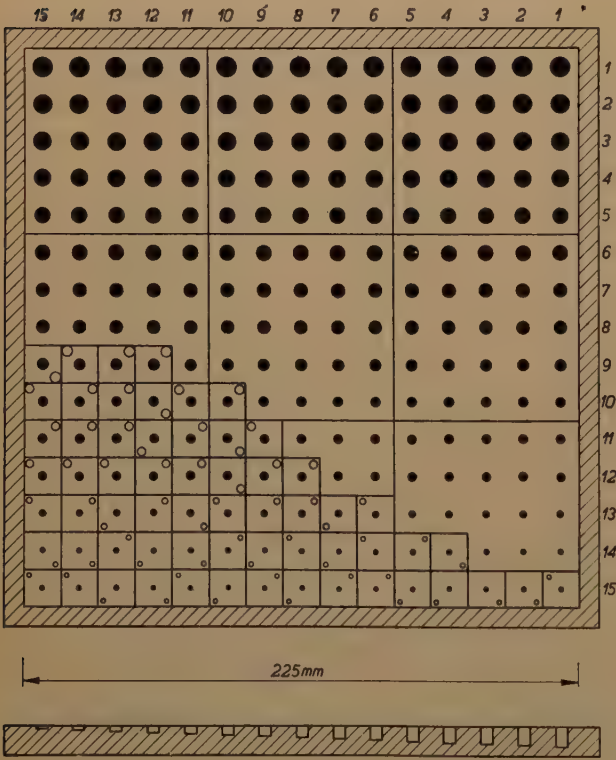
the attenuation of X-radiation. The following table gives a comparison between some of these properties for “Philite” and for water:

	“Philite”	Water
Average atomic number	6.9	6.0
Density	1.3 g/cm ³	1.08 g/cm ³
Ratio of transmission at 100 kV and at 50 kV	1.34	1.33
Mass absorption coefficient	2.3 λ ³ cm ² /g	2.6 λ ³ cm ² /g
Mass scatter coefficient	0.18 cm ² /g	0.19 cm ² /g

Thus it is seen that “Philite” does indeed correspond well to the human tissue, which for 70% consists of water. Moreover, it is an easily workable material.

A fluorogram of the phantom produced on a 45 mm film with an apparatus for mass chest radiography is reproduced in fig. 4. From this it is fairly simple to judge the effect of contrast and size of detail upon the visibility of the details, both separately and together.

Roughly speaking there are two areas to be distinguished in this radiograph. In the first area all the holes are visible, whereas in the other it is impossible to discern any detail at all. There is no sharply defined transition between these two areas. In order to determine the boundary as well as possible, in the part of the object plate where the holes are small in diameter and shallow in depth additional holes have been drilled (indicated in fig. 1 by circles); these have exactly the same dimensions as the corresponding holes situated in the centre of the respective squares marked off



60260

Fig. 1. Object plate for contrast-detail diagram. The holes in each vertical column are of the same depth but vary in diameter; those in a horizontal row have the same diameter but vary in depth. Near the small holes additional holes have been drilled which are indicated by circles.

on the phantom with lead oxide. The extra hole may be in any one of four possible positions in its square, and in order to decide whether a certain hole is visible or not all the observer has to do is to indicate the position in which he thinks the extra hole is situated. In this way the border line can be determined fairly sharply.

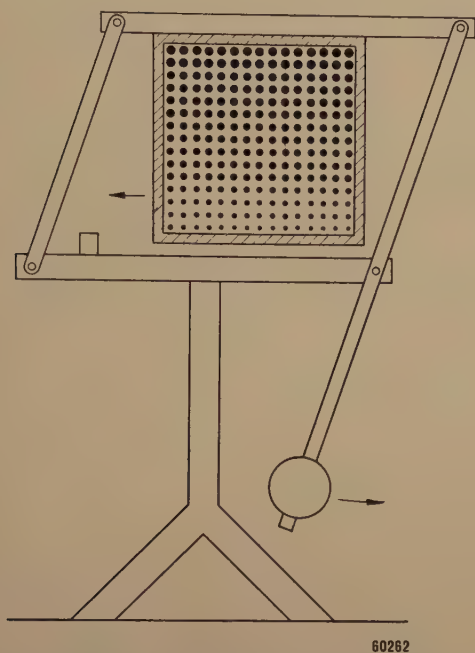


Fig. 3. Apparatus for determining the effect of movement of the phantom upon the sharpness and contrast of the X-ray image. The phantom is suspended in a frame made in the form of a hinged parallelogram. By means of a pendulum coupled to the frame an almost purely translational movement is imparted to the phantom. A device not shown in the drawing ensures that the motion is practically uniform, whilst the speed of the movement can easily be adjusted and recorded.

In order to construct a contrast-detail diagram (fig. 5) the diameters d of the holes are plotted along the ordinate and their depths h along the abscissa. The dimensions of the holes that are just visible in the radiograph correspond to a curve in the diagram, the visibility curve. And since the diagram represents, as it were, the phantom itself, the visibility curve corresponds to the border line on the phantom. This line has roughly the form of a hyperbola, which means to say that small details must have great contrast and details with a small contrast must have large dimensions if they are to be perceived. When we attempt to define the resolving power of the X-ray method employed as the reciprocal value of the volume of the smallest visible object (in the form of a cylinder) then it becomes apparent that this factor largely depends upon the shape of the cylinder to be observed; in other words, the curve of observation is not a curve of constant volume

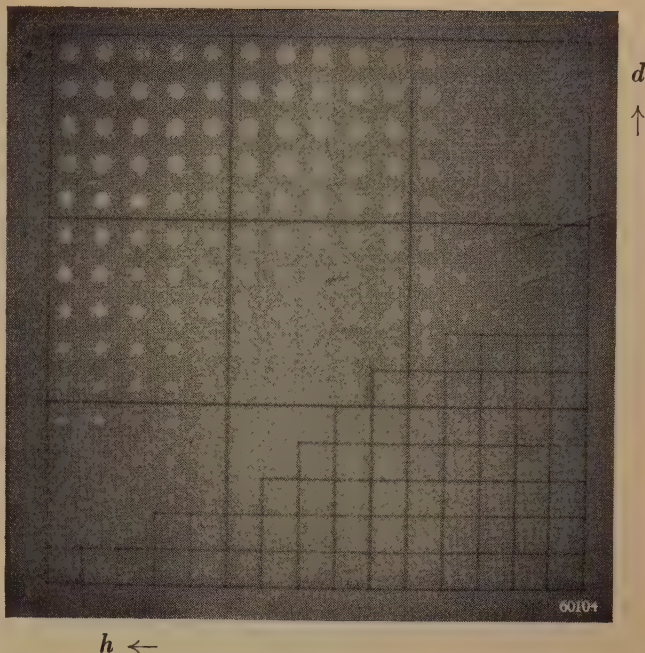


Fig. 4. Fluorogram of the contrast-detail phantom on 45 mm film. (The reproduction is the reflected image of fig. 1.)

(cf. the dotted curve in fig. 5). A thin and relatively long cylinder is more easily perceived than a short and thick cylinder, as is to be seen from fig. 6a, where the reciprocal of the detail has been determined and plotted against the diameter of the holes for various points along the curve of observation, and from fig. 6b, where the reciprocal volume has been plotted against the depth of the holes. What is of particular interest for prac-

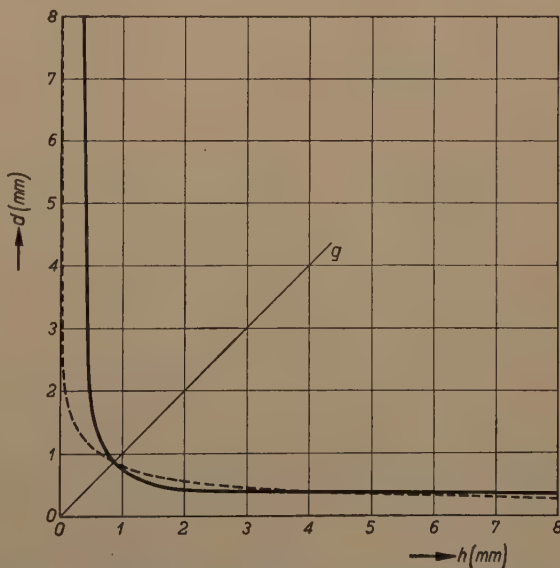


Fig. 5. Contrast-detail diagram of a contact radiograph. Full line: the visibility curve. Dotted line: a curve of constant volume; the objects on this curve all have the same volume as that of the smallest visible cylinder with diameter equal to height. The dimensions of this cylinder are indicated by the point of intersection of the observation curve with the line g drawn at an angle of 45° through the origin.

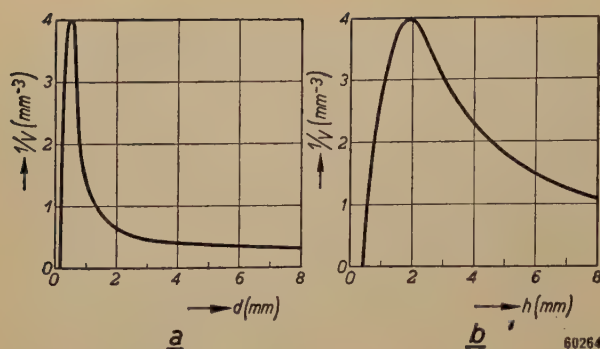


Fig. 6. a) Reciprocal volume V of the holes just visible according to the visibility curve of fig. 5, plotted as a function of the diameter of the holes. b) ditto, plotted as a function of the depth of the holes.

tical purposes is a cylinder of which the diameter and height are equal, since this most closely approaches a sphere. (A recently formed nodule in the lung is also roughly spherical.) The smallest still visible cylinder of this shape is given in the contrast-detail diagram by the point of intersection of the visibility curve with the straight line (g) drawn at an angle of 45 degrees through the origin (fig. 5). In a fluorogram like that in fig. 4 the size of this cylinder is about 2 mm.

Apperception diagram and the phantom used for it

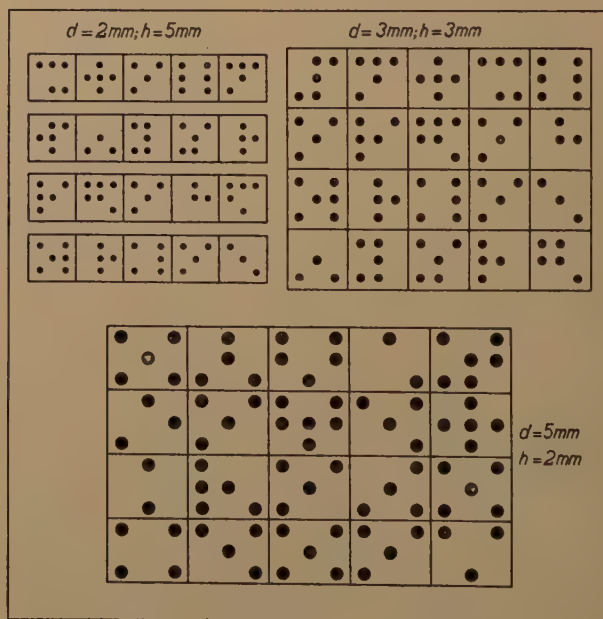
It appears that the time taken to judge an image of the phantom described in the foregoing depends largely upon the quality of the X-ray image. The better the X-ray method the better is the curve and the less time it moreover takes to observe it. The latter factor does not find expression in the contrast-detail diagram.

Furthermore, any judgment of an X-ray method according to what is revealed by the contrast-detail diagram does not fully answer practical requirements, inasmuch as the phantom test discloses whether details in the image can be observed at places where one knows they should really be, whereas in actual practice it is not known where the details sought may occur.

For this reason, following upon some tests that were carried out in regard to the influence of an enlargement of the X-ray image (see below), a so-called "apperception phantom" has been constructed which makes it possible to ascertain how long it takes to judge an X-ray image in its details and how many errors are made in that judgment.

The object plate of the apperception phantom (fig. 7) also has a number of holes, but not arranged in regular rows. The plate is divided into three parts, each marked off into twenty squares with

lead oxide. In each square is a number of holes varying from two to six, all in different positions. The holes in each part of twenty squares are of the same dimensions, viz. diameter 2 mm and depth 5 mm, 3 mm and 3 mm, and 5 mm and 2 mm respectively for the three parts. These dimensions have been so chosen that the detail falls well within the limit of visibility as was determined from the contrast-detail diagram for an apparatus for fluorography. The observer counts the number of holes he can see in a square in a radiograph or fluoroscopic image of the phantom, and as quickly as possible. If he fails to give the right number this counts as an error. The time taken to count the holes in twenty squares of one of the three parts (which part depends upon the wishes of the observer) is noted down⁵⁾. In a diagram constructed by plotted the time along the abscissa and the number of errors along the ordinate a complete observation corresponds to one point. Different X-ray methods will give different points, and in this way an "apperception diagram" is obtained, an example of which is reproduced in fig. 8.



60265

Fig. 7. Object plate for the apperception diagram. This is divided into three parts. The holes in each part have the same dimensions, respectively 2 mm diameter and 5 mm in depth, 3 and 3 mm, and 5 and 2 mm. The holes are drilled in groups, with two to six holes in each square marked off with lead oxide.

⁵⁾ A similar method for judging the perceptibility of details has also been applied for formulating illumination standards according to Weston, as described by A. A. Kruithof and A. M. Kruithof in their article: Basic principles for the formulation of illumination standards, Philips Techn. Rev. 10, 214-220, 1949 (No. 7).

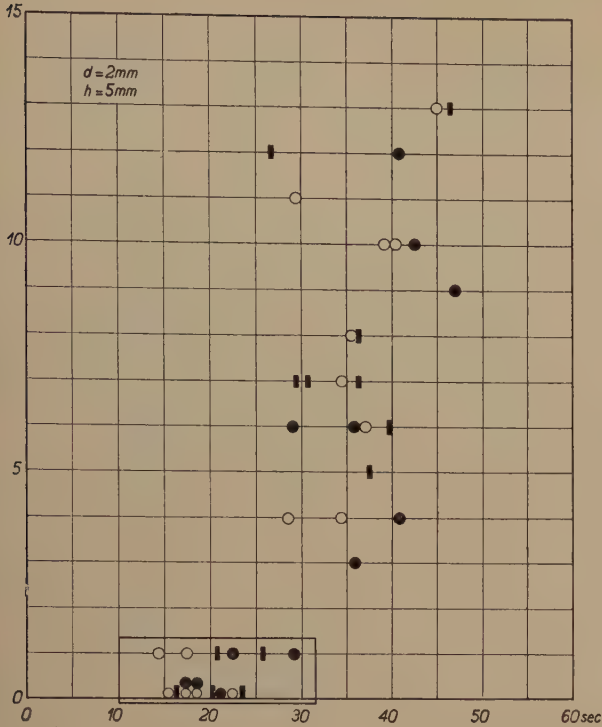


Fig. 8. Example of a apperception diagram. The observation time is plotted in seconds along the abscissa and the number of errors made by the observer along the ordinate. Each complete observation furnishes a point in the diagram. The points indicated by different signs refer to different films and different ways of observing the image of the apperception phantom produced by the method of fluorography (through a magnifying glass or by means of projection). The points indicated by the same sign refer to observations with films of different density. The frame at the bottom encompasses all the points obtained when judging enlarged radiographs.

Applications of the diagrams

a) With the aid of contrast-detail diagrams and apperception diagrams it can now be decided whether certain X-ray methods merit preference over others.

Owing to the aforementioned dependency of the resolving power upon the shape of the object an X-ray method is not to be judged solely, for instance, from the smallest visible sphere. Judging according to the shape of the whole of the contrast-detail curve is far more complete. The closer this curve follows the coordinate axes the better does the respective X-ray method answer the purpose. In eleven Scandinavian institutes the writer plotted contrast-detail diagrams with the aid of the twelve apparatus used there for fluorography. The phantom was placed immediately in front of the screen (except in one case, No. 11, where an enlarged image was produced); both mirror and lens cameras were used. In fig. 9 twelve contrast-detail diagrams are represented for radiographs with an exposure time corresponding to a chest thickness of about 18 cm. From these diagrams it is to be concluded that there is a considerable difference in performance between the apparatus employed. Although such a difference is already fairly well indicated by the dimensions of the smallest visible cylinder with equal depth and diameter (see the points of intersection of the curves with the line *g*), the shape of the curves obviously furnishes a much better distinction.

b) The suitability of physicians and technicians for radiosopic work depends upon their ability to discern small details. This can be judged by getting them to plot a contrast-detail diagram of an image of the phantom on the fluoroscopic screen. It will be seen that there are striking differences between the results obtained by the various observers (see fig. 10). In this way it is possible to select the best observers.

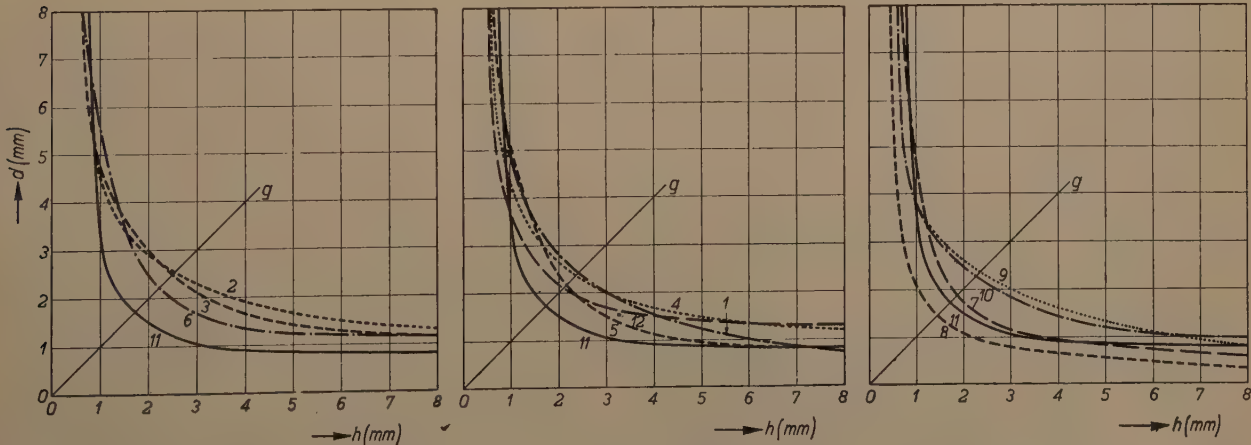


Fig. 9. Twelve contrast-detail diagrams recorded for comparing the properties of the X-ray installations used in eleven Scandinavian institutes. The dimensions of the smallest visible cylinder with diameter and height equal (intersection of the curves with the line *g*) already give an idea of the differences between the twelve methods; the shape of the observation curves as a whole makes these differences more pronounced. (To facilitate comparison the curve for method No. 11 is included in all three diagrams.)

c) In order to judge what improvement can be reached by X-raying with an enlarged image it is necessary to use, in addition to the contrast-detail diagram, also the apperception diagram,

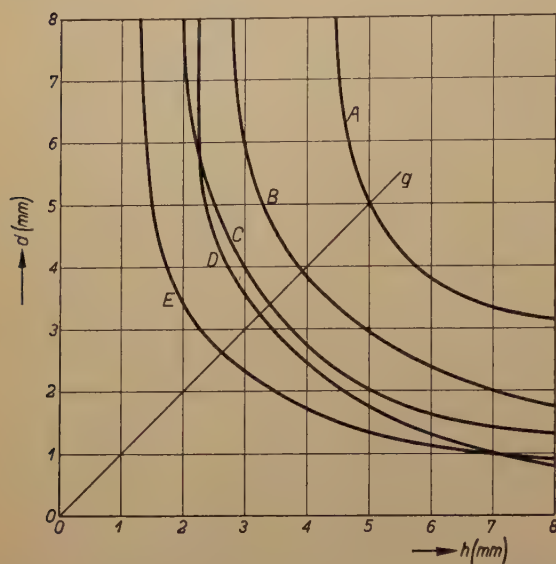


Fig. 10. Contrast-detail diagrams made by five different observers A-E from their examination of a fluoroscopic image of the contrast-detail phantom. The results of the five observers differ considerably.

because this diagram reveals best the difference between an enlarged image and a normal one. (It is in fact the investigation into this improvement that led to the designing of the apperception phantom.)

An enlarged image is obtained when the object is placed at a greater distance from the photographic film or screen than is normally the case (usually the object is placed as close as possible to the film or screen). The advantages derived from enlargement of the image are as follows (see footnote 4)):

- 1) The blurring due to the granular structure of the film or screen (intensifying screen) is of less importance.
- 2) Contrast is greater, because a larger portion of the scattered X-rays does not reach the screen.

However, there are also disadvantages attaching to an enlargement, viz:

- 1) Geometrical blurring is increased.
- 2) When screening it is easier to discover any abnormalities if the doctor can move the patient a little to and fro, but this cannot be done so well when the patient is some distance away from the screen.

In order to investigate the effect of an enlargement of the image a number of radiographs of both phantoms were made, both normal and enlarged

2.8 times, by the X-ray-camera method. The fluoroscopic image was photographed on a 45 mm film under different exposure conditions (different densities). The radiographs were examined in three different ways: under a magnifying glass with a power of 1.8, after projection on a white, fine-grained screen to a picture of 15 cm × 15 cm, and after projection to a picture of 35 cm × 35 cm. Contrast-detail diagrams and apperception diagrams were made of the pictures obtained. Both kinds of diagrams clearly show that the image quality can be considerably improved by enlargement, as may be seen from fig. 8 and fig. 11. The group of points within the frame at the bottom of fig. 8 relate to enlarged radiographs; they belong without exception to a small number of errors and a relatively short observation time. In fig. 11 the apperception curve relating to an enlarged image likewise shows a more satisfactory trend than that relating to the normal image. The improvement obtained by enlargement is expressed in figures in the table below. It appears that the relatively small improvement of 0.5 mm in the size of the smallest visible detail is accompanied by greater ease of apperception, as a result of which the time taken for the observation with the apperception phantom is reduced to about half and hardly any errors are made.

Image	Smallest visible detail	$d=5\text{ mm}, h=2\text{ mm}$		$d=2\text{ mm}, h=5\text{ mm}$	
		observ. time	errors	observ. time	errors
Normal	2.0 mm	49 sec	5.5	44 sec	7
Enlarged	1.5 mm	28 sec	1	19 sec	1

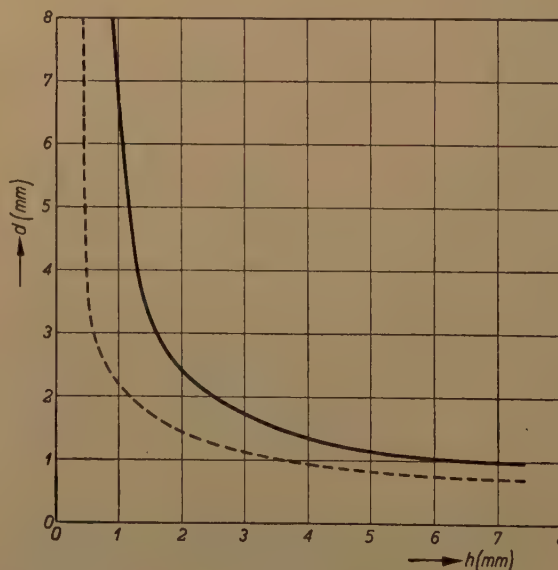


Fig. 11. Full line: contrast-detail diagram for screening with normal image. Dotted line: ditto with enlarged image.

The blurring of the screen, which sets a limit to the visibility of the details and is the main factor leading to the introduction of X-ray technique with enlarged image, can be determined by means of a so-called star⁶⁾. This star (see fig. 12) consists of a number of copper strips arranged as

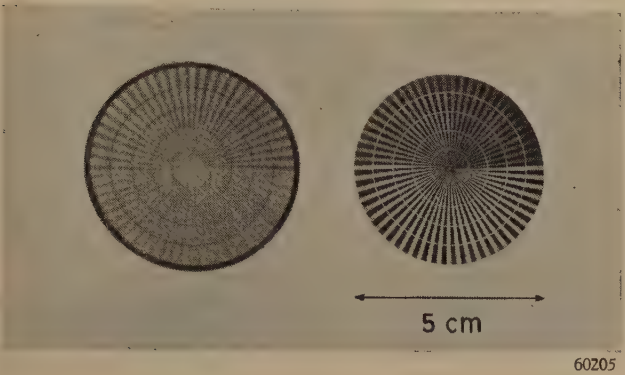


Fig. 12. Determination of the blurring of the fluorescent screen with the aid of a star consisting of a number of copper strips arranged as sectors in a circle. a) Photographic recording of the fluoroscopic image of the star. b) Contact radiograph of the star.

sectors in a circle. The space between the strips is filled with paper. The circle is divided into four quadrants, each with different thicknesses of the copper sectors, viz: 0.5, 1.0, 2.0 and 3.0 mm respectively. The radius of the circle is 2.5 cm. Four circular wires indicate where the distance between the copper strips is about 0.25, 0.50, 0.75 and 1.00 mm. Thus it is possible to determine with a fair degree of accuracy the smallest distance separating the sectors in the X-ray image. In order to determine the screen blurring the fluoroscopic image of the star is photographed together with a number of contact-radiographs of the star fixed onto the side of the screen facing the camera (see fig. 13). The difference in sharpness between the picture of the fluoroscopic image and that of the contact images (fig. 12) has to be ascribed entirely to the effect of the fluorescent screen (care has to be taken to ensure that the brightness and contrast of the screen image and the contact images are practically the same).

d) Finally mention has to be made of the possibility to investigate separately with the contrast-detail diagram the effect of the movement of

the object upon the properties of the X-ray image. It has already been pointed out that such a movement cannot be avoided when radiographing the lungs. The contrast-detail phantom is placed in the moving apparatus shown in fig. 3. Contact radiographs of the moving phantom with exposures of 0.05 sec and 0.25 sec give contrast-detail diagrams as illustrated in fig. 14. The explanation of the various curves is given in the subscript. A small movement is of little effect; although the object is less sharply defined it can still be observed, because the image is somewhat protracted by the movement. When, however, the motional unsharpness is greater than 2 mm visibility is noticeably reduced.

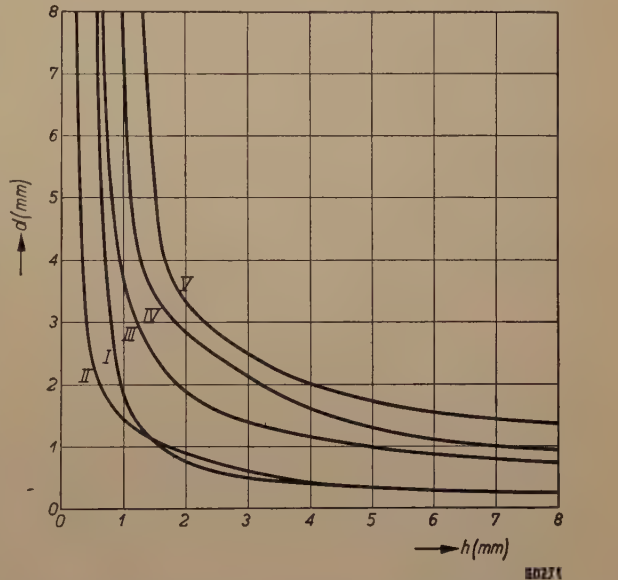


Fig. 14. Influence of the movement of the phantom upon the contrast-detail diagram. I motional unsharpness nil, II motional unsharpness 1 mm. Not much difference is to be seen between I and II. III, IV and V motional unsharpness respectively 2.5, 7 and 12 mm. The visibility curves show a less and less satisfactory trend.

Regular check of the quality of radiographs

In X-ray examination of the lungs it often happens that several radiographs have to be made of the same patient at more or less long intervals. To be able to compare these photographs and determine exactly how a certain affection of the lungs is progressing it is of the greatest importance to have data available regarding the sharpness and contrast properties of each picture. It is impracticable to collect these data with the aid of the contrast-detail phantom or the apperception phantom; not only are they too large to be photographed together with the patient, but it takes far too much time to scan the phantom images and construct the diagrams.

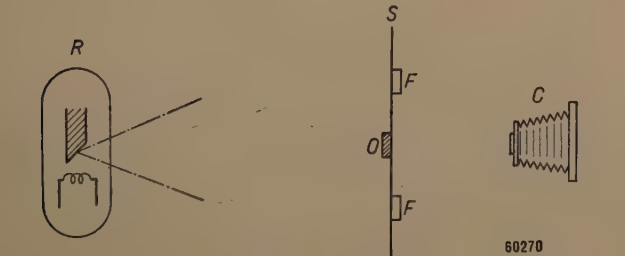


Fig. 13. Sketch of the set-up for determining the blurring of the fluorescent screen. R = X-ray tube, O = star, S = screen, F = contact radiographs of the star, fixed onto the screen, C = photographic camera.

⁶⁾ For another method of investigating this factor see H. A. Klasens, The blurring of X-ray images, Philips Techn. Rev. 9, 364-369, 1947/1948 (No. 12).

To form an idea of the properties of the thousands of lung radiographs that are made every year by the Medical Department of Philips Works at Eindhoven, a small phantom was designed which

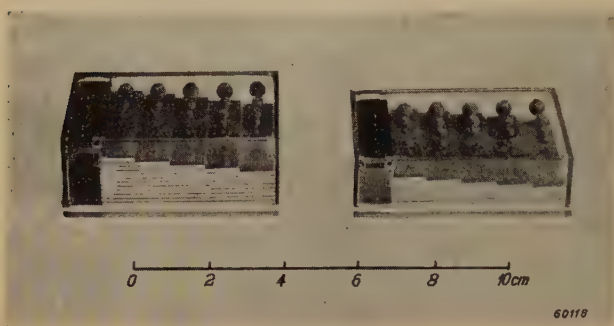


Fig. 15. Phantom for regular checking of the quality of radiographs. It is in two parts, each formed by five aluminium steps on which "Philite" spheres of varying diameter are affixed. At the left-hand side of the part on the left is a copper step and a strip of lead on the part on the right, for checking respectively the ray quality and the fogging of the film.

can easily be photographed together with the patient on either side of the neck and which enables the doctor to check regularly the quality of all his radiographs. This phantom (see *fig. 15*) consists of a set of aluminium "steps" cut into two parts. The first part has five steps of 4 to 8 mm thickness, the second part five steps of 9 to 13 mm. "Philite" cannot be used for this; owing to the small dimensions of the phantom "Philite" has too little total absorption for X-rays. For such small objects aluminium has been chosen as being a more suitable material. This, however, does not correspond so well to the physical qualities of the thorax, and consequently preliminary tests were carried out to compare an aluminium phantom with one of "Philite"; the former has been calibrated, as it were, with the aid of the latter. To represent the "details" five small spheres of "Philite" varying in diameter between 1 and 4 mm have been affixed to each step. Furthermore, to check the quality of the rays one part of the phantom is fitted with a step of copper (0.2 to 0.6 mm thick), whilst the other part has a strip of lead for checking the fogging of the unexposed film.

With the aid of this phantom it is quite easy to judge a radiograph of the lungs. In *fig. 16* a normal lung

radiograph is reproduced; the image of the phantom is to be seen above the patient's shoulders. The part of the photograph to be examined is compared with the aluminium step showing the same density, and the smallest detail of that step that can still be perceived is noted. It appears that in the case of contact radiographs on an average four and sometimes all five spheres can be observed on the aluminium step corresponding in density to the free field of the lung (between the ribs). This agrees well with the dimension of about 1 mm of the smallest visible cylinder on a contact radiograph as determined by means of the contrast-detail diagram.

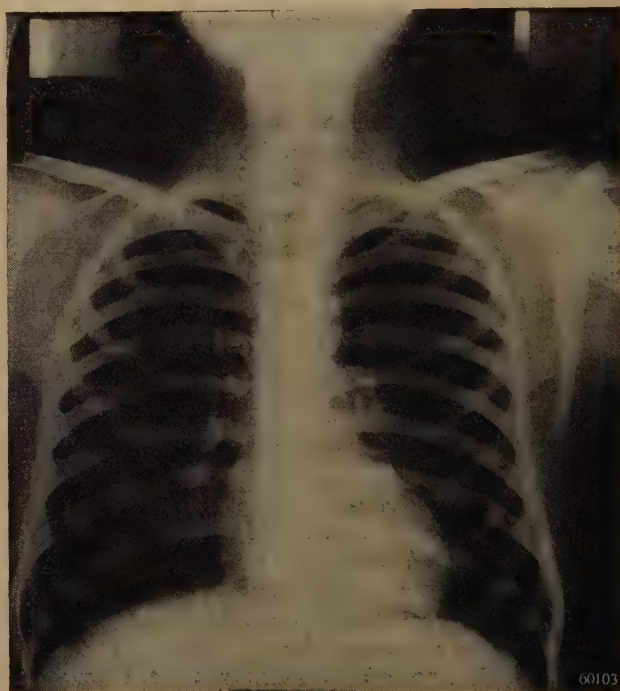


Fig. 16. Lung photograph, showing the check phantom above the patient's shoulders.

Summary. Following an introduction concerning the factors affecting an X-ray image, the construction is discussed of some phantoms enabling the merits of various radiologic methods to be compared. Some applications of these phantoms are dealt with: comparison of the properties of X-ray apparatus in different institutions, selection of doctors and others for X-ray screening, investigation into the influence of enlargement of the X-ray image, and the effect of movement of the object upon the X-ray image. Finally a phantom of smaller dimensions is described which serves for the regular checking of lung radiographs, which is particularly of importance when several radiographs have to be made of the same patient.

LIGHT SOURCES FOR LINE SPECTRA

by W. ELENBAAS and J. RIEMENS.

621.327.3/4:535.333.3

This article contains data appertaining to a series of discharge lamps which have been developed to serve as monochromatic light sources for optical experiments.

Introduction

In physcial experiments or tests where the light plays a part, and also in chemistry (for example in polarimetry), the need is often felt for a strong monochromatic light source, or for a light source which emits a number of monochromatic rays of a known wavelength. Sometimes one can manage with a "monochromator", an instrument in which a narrow part of the spectrum of a light source with a continuous spectrum (incandescent lamp or carbon arc) is separated with the aid of prisms and slits. In many cases, however, such a light source is not sufficiently monochromatic. One can then have recourse to the colouring of flames with metals, present in salts, as already applied by Kirchhoff and Bunsen. An example of this is the well known sodium flame used in polarimetry and for interference experiments. These flames, however, have the drawback that they are of low luminosity, whilst moreover the constancy of the radiation leaves much to be desired. A complicated apparatus then has to be devised for feeding the flame regularly with air containing a constant proportion of salt particles.

Shortly after the discovery made by Kirchhoff and Bunsen investigators in the field of spectroscopy began to make use of the fact that gases and vapours can be made to give light by means of an electric discharge. Plücker got his glass-blower Geissler to make some discharge tubes suitable for this purpose, which since then have been named after their maker. A Geissler tube (*fig. 1*) is in



Fig. 1. Geissler tube for use in spectroscopy.

essence a discharge tube with cold electrodes. The luminous part, called the positive column, is situated in a constricted part, so that the density of the current and the brightness are increased. Originally Geissler tubes were usually fed with the pulsating current from an induction coil. Provided care is taken to ensure adequate current

limitation they can also be fed with alternating or direct current of a high voltage. But there is always the drawback that the capillary and the electrodes become hot only under a relatively small load, as a result of which the gas becomes contaminated and the tube is rendered useless.

The development of modern gas-discharge lamps with oxide-coated cathode has created new possibilities, including such uses as referred to above, because an oxide-coated cathode can emit a strong thermionic current without being overloaded. It is now easy to make small discharge tubes containing some gas or other or a metallic vapour or a mixture of both in a very pure state and permitting a high current intensity, so that a light source is obtained capable of emitting considerable energy in one single spectral line or in a few lines.

Construction of the lamps

The discharge tubes at present being made by Philips for the above-mentioned purpose are either of normal glass or of quartz as needed. The dimensions of the discharge tube proper vary from case to case. The length of the luminous part of the discharge amounts to 3 to 4 cm.

At each end of the tube is a leading-in wire with an oxide-coated cathode attached, which is not separately heated but is brought to the required high temperature by the discharge and maintained at that level. In some cases, that is to say with some tubes filled with metallic vapour, the extremities of the discharge tube are metallized, with the object of keeping the temperature in the tube high enough to bring about a sufficiently high vapour pressure of the metal (e.g. cadmium). In *fig. 2* some forms of discharge tubes are shown diagrammatically in cross section.

In order to protect the discharge tube proper and to prevent oxidation of the leading-in wires, which sometimes get rather hot, the tube is built into an envelope, which is likewise made either of normal glass or of quartz as required. The space between the actual discharge tube and the envelope is evacuated or filled with nitrogen. The envelope is fitted with a screw base so that the lamp can be

inserted in a normal lamp-holder. A number of these lamps are shown in the photograph reproduced in *fig. 3*. The distance between the light centre and the lamp base is the same in all lamps, so that there is no need to readjust the height when a lamp has to be replaced.

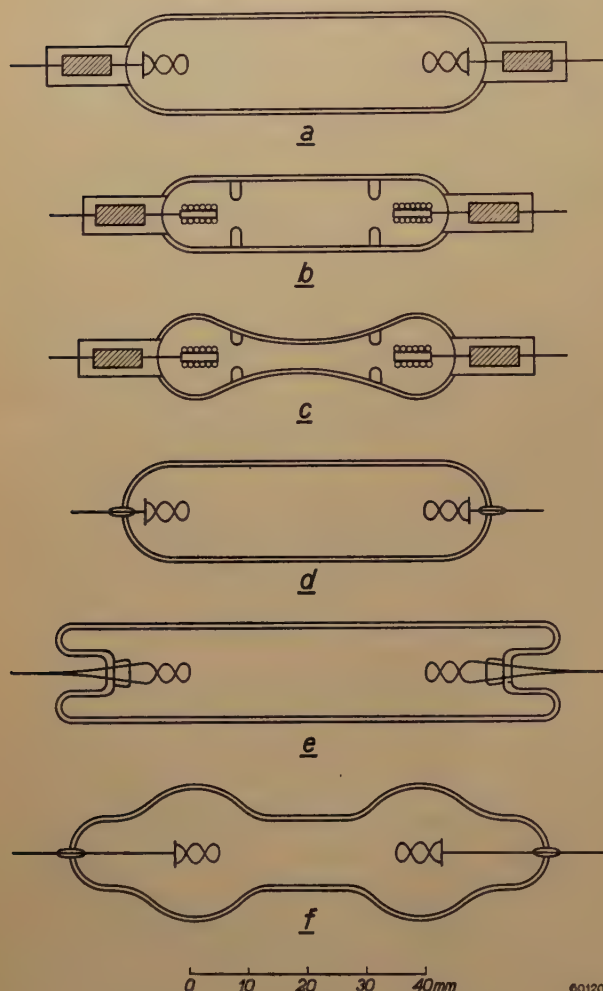


Fig. 2. Construction of the light sources for line spectra.
a) Mercury (low pressure, quartz),
b) mercury (high pressure) and cadmium (both in quartz),
c) zinc (quartz),
d) sodium, rubidium, cesium (glass),
e) mercury (low pressure), Ne, Ar, Kr, Xe (glass),
f) helium (glass).

In *b* and *c* (discharge tubes with high pressure) the oxide-coated cathode consists of a pin surrounded by coiled wire; in the other tubes the oxide-coated cathode is a twisted loop of coiled wire. In the quartz tubes *a*, *b* and *c* the inlet leads are foils of molybdenum. The tube *d* is lined with an alkali-vapour-resistant layer. The exceptional shape of *f* is necessary on account of the greater heat developed in helium.

Supply

Although some of these lamps could be connected to a 220 V A.C. supply, with of course suitable current limitation, for easy ignition it is better to supply them with a higher voltage. A transformer

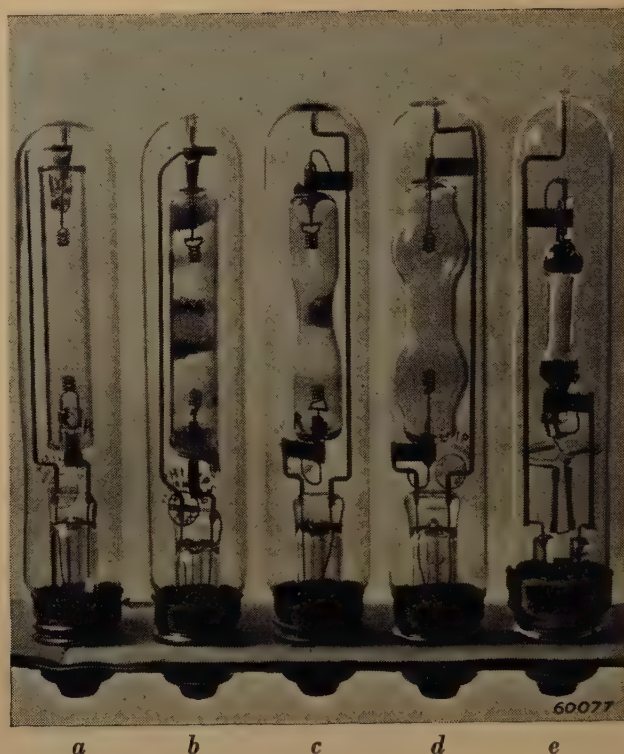


Fig. 3. Some lamps for line spectra.

- a) Mercury (low pressure, glass),
b) sodium,
c) rubidium,
d) helium,
e) zinc.

The metal condensed on the glass wall of the lamps *b* and *c* is vaporized by the heat developed when the lamp is burning.

having an open voltage of 470 V at a primary voltage of 220 V is therefore supplied together with the lamp. The transformer is an autotransformer of the low power factor type. This means that provision is made for a considerable spread of the lines of force, which has the same effect as if a choke were connected in series with a normal transformer. The current is thereby automatically limited. Consequently the working current is practically equal to the short-circuit current, which amounts to about 0.9 A. Thanks to the high open voltage the tube ignites without the oxide-coated cathodes having to be pre-heated. The circuit is diagrammatically represented in *fig. 4*.

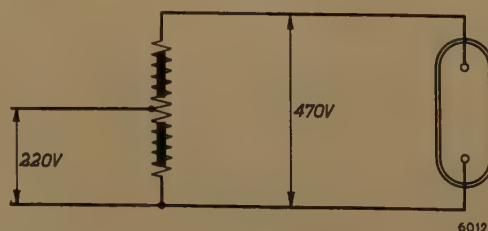


Fig. 4. Circuit diagram of the supply apparatus (autotransformer with low power factor) with the lamp connected to it.

The spectra

The lamps are filled either with one of the rare gases He, Ne, A, Kr, Xe, or with argon to which a metal has been added, this metal being vaporized by the heat of the discharge. Suitable metals are the alkali metals Na, Rb, Cs and the bivalent metals Zn, Cd, Hg. Mercury lamps can be made

for low as well as for high pressures, in the latter case the amount of mercury being such that the metal is entirely vaporized at the working temperature. Fig. 5 is a reproduction of photographs of the visible spectrum of the various lamps, whilst in fig. 6 the ultra-violet spectra are shown. If it is desired to separate a part of the spectrum,

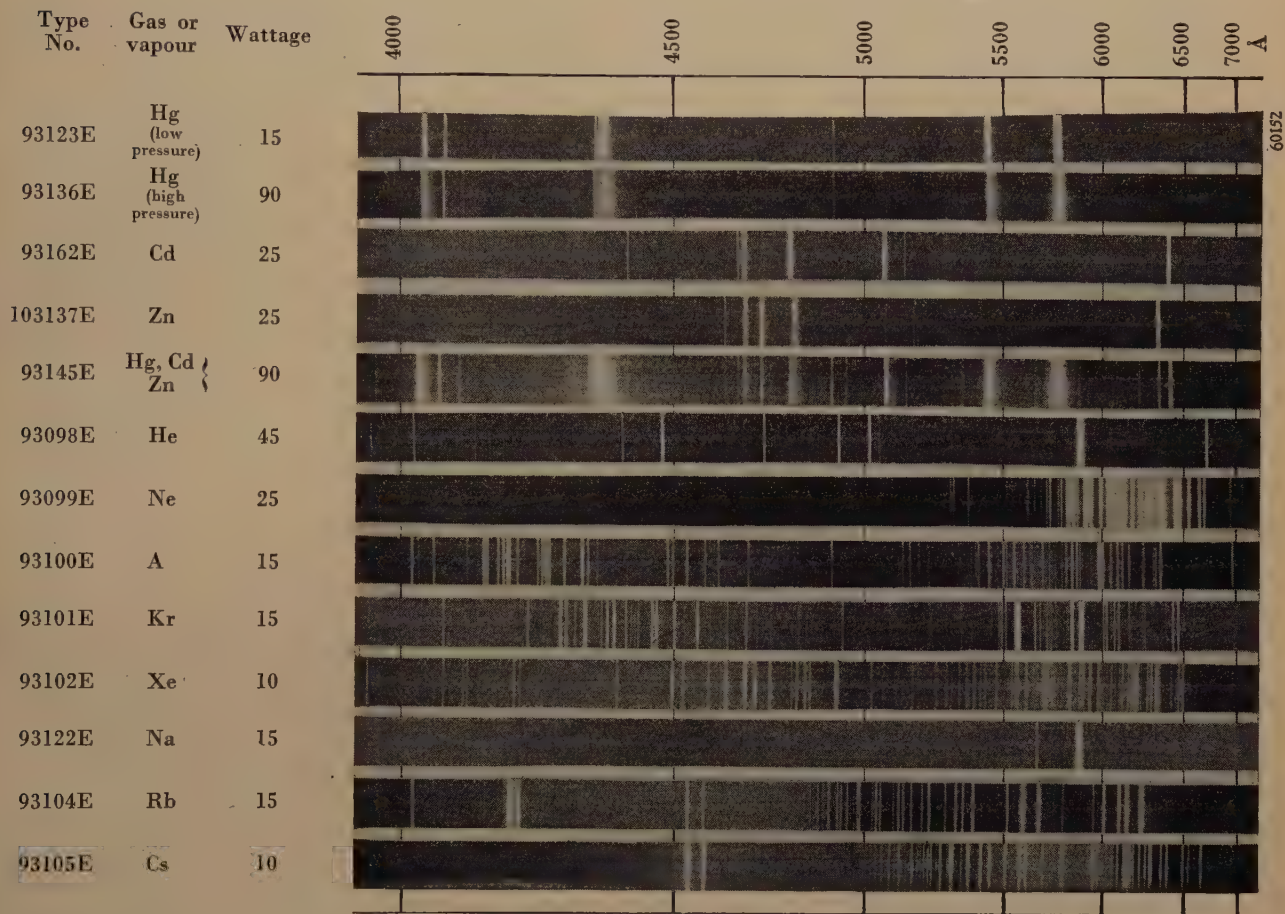


Fig. 5. Spectra of the various types of lamps that can be used for the visible spectrum. The envelope of all these lamps is of glass. The type number, the gas or vapour filling and the wattage of the lamps are indicated on the left.

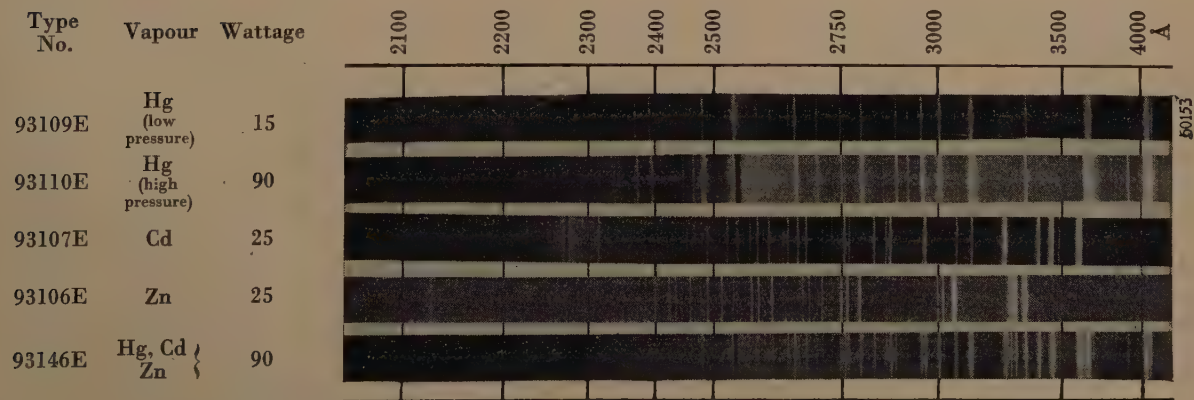


Fig. 6. The same as fig. 5 for the ultra-violet spectrum. All these lamps have an envelope made of quartz.

Table I. Combinations of lamp types and filters with which certain spectral lines can be isolated. If it is necessary to eliminate the infra-red radiation then the filters indicated in the last column have also to be added. The filters A, B, C, D and E are liquid filters of the composition given in table II (to be used in an optical glass cell 20 mm thick). The others, with the exception of those shown as Zeiss filters, are Schott filters. (The data for this table have been taken mainly from: H. Ewest and K. Larché, *Handbuch der Lichttechnik*, Springer Berlin, 1938, p. 192.)

Isolated wavelength in Å	Type No. of lamp	Filters	Filters for absorption of infra-red
3076	93106E	UG 5 (5 mm) + A + B (10 mm)	E
3126/32	93110E	UG 5 (5 mm) + C	
3261	93107E	UG 2 (2 mm) + A + C	
3341	93110E	UG 2 (2 mm) + A + D	E + BG 19 (2 mm)
3650/63	93110E	UG 2 (2 mm) + BG 12 (4 mm)	
4047	93123E 93136E	UG 3 (9 mm) + GG 4 (1.5 mm)	
4358	93123E 93136E	BG 12 (4 mm) + GG 3 (4 mm) or Zeiss C	E + BG 19 (2 mm)
4555/93	93105E	GG 2 (2 mm) + BG 12 (2 mm)	E + BG 19 (2 mm)
4678/4800	93162E	GG 5 (1 mm) + BG 12 (2 mm)	E + BG 19 (2 mm)
5086	93162E	GG 11 (2 mm) + VG 3 (2 mm)	E + BG 19 (2 mm)
5461	93123E 93136E	BG 11 (20 mm) + OG 1 (1 mm) + BG 18 (3 mm) or Zeiss B	E + BG 19 (2 mm)
5770/91	93123E 93136E	OG 3 (1 mm) + VG 3 (1 mm) + BG 18 (1 mm) or Zeiss A	E + BG 19 (2 mm)
5890/96	93122E	OG 3 (1 mm) + VG 3 (1 mm) or OG 2 (2 mm)	E + BG 19 (2 mm)
6362	103137E	RG 1 (2 mm)	BG 19 (2 mm)
6438	93162E	RG 1 (2 mm)	BG 19 (2 mm)

Table II. Liquid filters used for isolating certain spectral lines according to table I.

No.	Composition	Weight per litre water
A	Nickel-cobalt sulphate $\text{NiSO}_4 \cdot 1 \text{ H}_2\text{O}$ $\text{CoSO}_4 \cdot 1 \text{ H}_2\text{O}$	303 g 86.5 g
B	Picric acid	31.6 mg
C	Potassium chromate K_2CrO_4	150 mg
D	Nitric acid HNO_3	12.6 g (0.2 n)
E	Copper sulphate $\text{CuSO}_4 \cdot 5 \text{ H}_2\text{O}$	28.5 g

filters can be used. In favourable cases it can be so arranged that only light of one wavelength is emitted. If under certain circumstances it is not possible to achieve this with filters then a monochromator will have to be placed behind the lamp. In *table I* a number of lines are indicated which can be isolated with a suitable combination of lamp and filters; *table II* gives the composition of the liquid filters occurring in the combinations.

Summary. A description is given of the construction and supply apparatus of a number of light sources for line spectra, as also a summary of the various fillings (rare gases and metallic vapours) and the spectra emitted. A table shows how certain rays can be isolated from the various spectra with the aid of certain filters, which are further described

INVESTIGATIONS INTO THE IMPACT STRENGTH OF IRON AND STEEL

by J. D. FAST.

620.178.746.22:669.12:
669.141.24

For judging the strength of steels the so-called notched-bar impact test is often applied in addition to the usual tensile test. This impact test makes heavier demands upon the resistance of the material to brittle fracture than the tensile test, which mainly gives indications of the resistance to plastic deformation. This article deals not only with matters of common knowledge such as the execution and significance of the notched-bar impact test and the elementary concepts of stress and fracture, but also with original measurements by means of which the impact value of pure iron is determined as a function of temperature, and the effect thereon of oxygen, carbon and nitrogen is investigated.

The notched-bar impact test

In exceptional circumstances, steels which are shown to be ductile by the normal tensile test (great deformation preceding fracturing), when in actual use break in a "brittle" manner (no previous deformation worth mentioning). In the past ten years or so the collapse of some welded bridges and several welded ships has caused considerable consternation. The fractures found in these cases appeared to be of the brittle type, notwithstanding the fact that under tensile test the materials had behaved reasonably well. These and other experiences had long ago led to the view that the question whether fracture is tough or brittle is determined not only by the properties of the material but also by the manner in which the material is loaded during testing or in use.

The important part that the manner of loading plays in tough or brittle fracture is evident from bending tests with notched bars. Long ago it was found that such tests make heavier demands upon the resistance of materials against rupture than the classical tensile test. The notched bars were bent in a vice in such a way that the notch opened while the material was being deformed. It was occasionally found that, using two kinds of steel which behaved in practically the same way during tensile testing, one showed tough fracturing in the bending test, whereas the other showed brittle fracturing. The test could be made more severe still by giving the clamped bars a blow with a hammer over the notch so as to cause them to bend more quickly. Based upon these findings machines have been constructed which enable us to test the metal in such a way that a satisfactory result of the test (notched bar impact test) offers a better guarantee against accidents than a satisfactory result of the normal tensile test.

Various types of such machines are in use in different countries and the specifications for the

dimensions of the bars to be tested also differ widely. In our investigations we used bars of 10 mm × 10 mm × 55 mm with a notch 2 mm wide and 5 mm deep (see fig. 1), as recommended by the International Federation of the National Standardizing Associations, and tested them in a

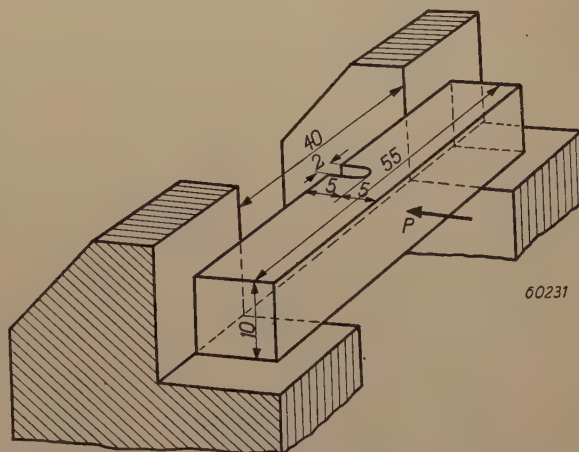


Fig. 1. Sketch representing a notched bar for the impact test. The dimensions given are in mm. The bar is struck by the hammer at the point P .

Charpy machine. In this test the bar rests on two horizontal supports and against two vertical supports 40 mm apart and is struck in the centre of the plane opposite the notched side by a pendulum hammer H released from a certain angle α corresponding to the height h (see fig. 2). The hammer is attached to an arm S . The shape and distribution of the mass of the hammer are so chosen that when the blow is struck practically no shock is transmitted to the pivot of the arm¹⁾. Upon the blow being struck the notched bar breaks or is torn open and bent so far that it can be forced out

¹⁾ For this purpose the hammer has to strike the bar with its centre of percussion not far removed from its centre of gravity.

through the opening of 40 mm between the supports. Under the action of its remaining energy the hammer then rises on the other side and reaches, say, an angle β corresponding to the height h' . The work A performed upon the bar is given by the equation

$$A = G(h - h') = Gl(\cos \beta - \cos \alpha), \dots (1)$$

where G is the weight of the hammer (in our experiments 10 or 30 kg, or 98 and 294 newtons respectively) and l is the distance between the pivot and the centre of gravity of the pendulum.

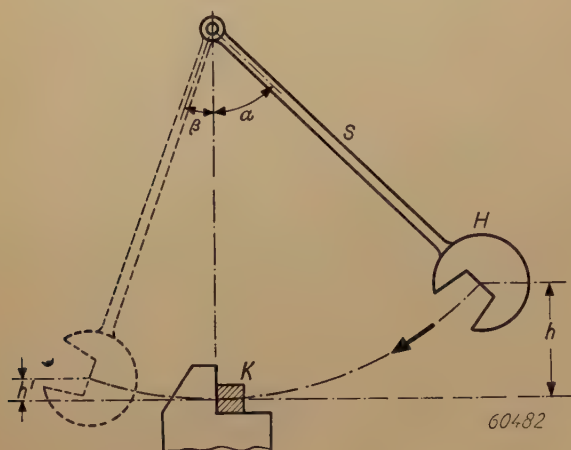


Fig. 2. Diagrammatic representation of the Charpy machine for determining the impact value. The pendulum hammer H is so constructed that its so-called centre of percussion coincides with the point where it strikes the notched bar K . Given the angle α , the work performed by the pendulum hammer can be determined by measuring the angle β .

The impact work A (in kgm), converted per cm^2 of the plane of fracture, in our case for a plane of $10 \text{ mm} \times 5 \text{ mm}$, multiplied by two, is called the notched-bar impact strength (or impact value). This value has no other direct physical significance but gives a good technical measure of the resistance of the material against fracturing. If the break is preceded by an appreciable plastic deformation the impact value is high and the metal usually shows a "fibrous" ("mat") fracture. If there is practically no plastic deformation then the impact value is low and as a rule the metal shows a "granular" ("bright") fracture. Between these two cases are "mixed fractures", having partly a "fibrous" and partly a "granular" appearance.

Significance of the notched-bar impact test

The phenomenon that one and the same material at the same temperature shows ductility under one test (e.g. the tensile test) and brittleness under another (e.g. the impact test) is mainly due to the fact that under these tests the material is subjected to different states of stress. (The

difference in the rate at which the test is carried out is of less importance; see also pages 307-308.) The state of stress in any particular point of a body can be described by imagining a large number of planes drawn through that point and indicating the magnitudes and directions of the forces acting upon these differently orientated planes per unit area. Each of these stresses can be represented by a vector. This vector can be resolved into a component lying in the particular plane, the shear stress, and a component perpendicular to that plane, called the normal stress. It can be proved that under any load (however complicated) in each point of a body three mutually perpendicular planes can be indicated in which the shear stresses are nil. The normal stresses on these planes are called the principal stresses. Once the magnitudes and the directions of these three principal stresses are known, then the state of stress in the particular point is fully known, for the stresses on all the plane elements differently orientated through that point can be found by calculation. These three principal stresses will be denoted by the symbols σ_1 , σ_2 and σ_3 , the first one denoting the greatest and the last one the smallest (in the algebraical sense). Tensile stresses are reckoned as being positive and compressive stresses negative, so that a compressive stress is always smaller than a tensile stress. The principal stress σ_1 is also the greatest of the normal stresses acting in the particular point:

$$\sigma_{\max} = \sigma_1. \dots \dots \dots (2)$$

The maximum shear stress is given by:

$$\tau_{\max} = \frac{1}{2}(\sigma_1 - \sigma_3) \dots \dots \dots (3)$$

It acts in each of the two mutually perpendicular planes bisecting the right angles between the directions of σ_1 and σ_3 and containing the direction of the principal stress σ_2 . In the ordinary tensile test ($\sigma_2 = \sigma_3 = 0$) the maximum shear stress acts in all planes making an angle of 45° with the direction of the tensile force.

The normal stresses tend to tear the body apart along the planes on which they are acting, whereas the shear stresses tend to shear the adjacent planes of the body along each other, thus being responsible for plastic deformations, where these occur. To be able to predict whether a body can bear a certain state of stress without plastic deforming and without fracturing, we must know its resistance to shear and its resistance to cleavage.

The significance of these two factors is most readily seen in the case of single crystals: when

the shear stress acting in a certain crystallographic plane and in a certain direction exceeds a critical value (the resistance to shear, or shear strength, for that plane and that direction), then permanent deformation takes place owing to the planes shearing one over the other. When the normal stress acting perpendicular to a certain crystallographic plane exceeds another critical value (the resistance to cleavage, or cleavage strength, for that plane) then fracturing takes place.

It is obvious that even in the case of a single crystal there is already a complexity of shear and cleavage strengths to be distinguished. Still more complicated, at least in principle, is the situation in polycrystalline metals, with which one always has to deal in technical engineering. If, however, the crystals are rather small and contain no preferred orientations then certain mean properties can be ascribed to the polycrystalline material, thus one particular shear strength τ_{cr} and one particular cleavage strength σ_{cr} independent of position and direction. These concepts now have the significance of certain critical values of the maximum shear stress τ_{max} and maximum normal stress σ_{max} given by the load. Plastic flow begins to take place as soon as

$$\tau_{max} > \tau_{cr}, \dots \dots \dots (4)$$

and rupturing takes place as soon as

$$\sigma_{max} > \sigma_{cr} \dots \dots \dots (5)$$

When, as the load is increased, σ_{cr} is exceeded before τ_{cr} then the material breaks without any previous plastic deformation taking place. If, on the other hand, τ_{cr} is first exceeded then fracturing is preceded by deformation, to a greater or lesser extent.

According to a criterion which in most of the cases investigated proves to be somewhat more accurate than that of the maximum shear stress τ_{max} , flow begins in an isotropic polycrystalline metal as soon as a quantity taking the place of τ_{max} exceeds a critical value depending not only upon σ_1 and σ_3 , as is the case with τ_{max} (see formula (3)), but also upon the principal stress σ_2 . For our considerations, which are for the greater part qualitative, however, nothing is to be gained by going into this point any farther; neither shall we go farther into the fact that in certain cases the occurrence of fracture is determined by a criterion other than that of the maximum normal stress.

Now in order to understand how it is possible that in certain cases brittle fracture may occur with the impact test and not during the ordinary tensile test, we shall first consider the extreme case (impossible of practical realization) of a

body subjected to a uniform triaxial tension ($\sigma_1 = \sigma_2 = \sigma_3$). According to formula (3) no shear stresses then arise in any of the intersecting planes of the body. Under such a loading with continuously increasing tensile stresses all bodies would break without any plastic deformation previously taking place, even bodies of a material that would appear to be highly ductile in a normal tensile test.

During the impact test the three principal stresses in the material underneath the notch are not, it is true, mutually equal, but when the loading is applied then there arises in that part of the material a considerable triaxial tensile stress, in contradistinction to the uniaxial stress arising in the normal tensile test. This can be explained in the following way.

The material adjacent to and lying underneath the base of the notch is locally elongated in the longitudinal direction of the test bar on account of a large tensile stress arising there as the bar is bent. This is accompanied by a contraction (Poisson contraction) in all directions perpendicular to the direction of the tensile stress. The notch, however, forms an interruption in the continuity of the bar, so that the material adjacent to the sides of the notch does not come under load and does not take part in the deformation. This forms a hindrance for the Poisson contraction in the deformation zone and results in tensile stresses perpendicular to the direction of the primary tensile stress. According to formula (3) the occurrence of these secondary tensile stresses implies that the maximum normal stress $\sigma_{max} (= \sigma_1)$ arising in the impact test must be greater than that arising in the ordinary tensile test in order to get the same value of the maximum shear stress τ_{max} .

In the ordinary tensile test σ_{max} is twice as great as τ_{max} under any loading. If the load on the test bar is gradually increased from zero, σ_{max} and τ_{max} increase (without their ratio $\sigma_{max}/\tau_{max} = 2$ being changed), and it will depend entirely upon the values of the cleavage strength σ_{cr} and the shear strength τ_{cr} whether the material will ultimately fracture in a brittle way or whether a larger or smaller plastic deformation will precede the rupture. It goes without saying that the latter will be the case when $\sigma_{cr}/\tau_{cr} > 2$.

According to the foregoing, in the impact test $\sigma_{max}/\tau_{max} > 2$. Assuming, for the sake of argument, that in the case of the notched bar used by us $\sigma_{max}/\tau_{max} = 3$, then if it is to show ductility it will be necessary that $\sigma_{cr}/\tau_{cr} > 3$,

which means to say that the cleavage strength of the material of the bar must be more than three times its shear strength. If in a certain material σ_{cr}/τ_{cr} were to have the value 2.5 then that material would show ductility in the ordinary tensile test and brittleness in the impact test; the absorption of energy in the tensile test would then be relatively great but in the impact test relatively small (low impact value), because comparatively little energy is required to bring about the fracture itself.

If for a certain material subjected to the impact test the relation $\sigma_{cr}/\tau_{cr} > \sigma_{max}/\tau_{max}$ is satisfied, then there is still no certainty that a high impact value will be found, or, in other words, that considerable deformation will precede the fracture, since from what follows later it will be seen that σ_{cr}/τ_{cr} decreases as the deformation increases. If, therefore, at the beginning of the test this ratio is only little greater than σ_{max}/τ_{max} then fracturing will take place after comparatively little deformation, so that the impact value will be low, though higher than it would be in the case of fully brittle fracturing. For a metal like iron the quotient σ_{cr}/τ_{cr} also varies with the temperature, so that it depends to a high degree upon the temperature at which the test is carried out whether brittle fracturing will occur or not. We shall defer the discussion of this influence of temperature till later, when dealing with the experiments that we carried out.

Influence of oxygen and carbon on the impact value of iron at 20 °C

Much technical research work has already been done with regard to the impact value of iron and steel, but the influence of each impurity and each component of an alloy separately has not yet been investigated, nor even has the impact value of pure or practically pure iron ever been measured as a function of temperature. In the following pages, therefore, results will be given of some experiments relating thereto. The preparation of the iron and the various alloys needed for these experiments has already been described in this journal²⁾.

The impact value of pure iron (carbon and nitrogen contents less than 0.001 wt %, oxygen content about 0.001%) at 20 °C was found to be 17 kgm/cm². Relatively small additions of oxygen were sufficient to reduce considerably the impact value at that temperature, as demonstrated by the lower curve in fig. 3, which shows, for instance, that an oxygen content of 0.018% caused

the impact value to drop to 3.3 kgm/cm². On the other hand, such an oxygen content has but very little effect upon the behaviour of iron during the normal tensile test.

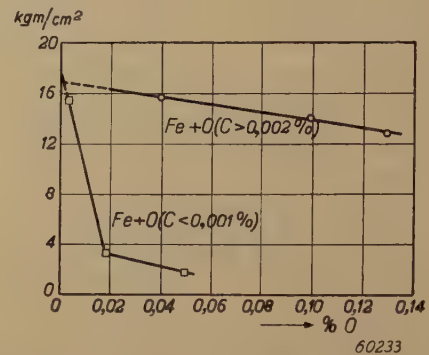


Fig. 3. Influence of the oxygen content of iron upon the impact value. When the iron contains less than 0.001 wt % carbon the impact value drops to low values upon the addition of only a few hundredths wt % of oxygen (lower curve). If, however, the iron contains a small quantity of carbon (e.g. 0.002%) the influence of the oxygen disappears almost entirely (upper curve). For those who prefer to use the Giorgi units, it is to be noted that 1 kgm/cm² $\approx$ 10⁵ Nm/m².

The great influence of oxygen upon the impact value of iron points, according to what has been stated in the foregoing, to a lowering of the quotient σ_{cr}/τ_{cr} by oxygen. In many cases the fracture of impact-test bars containing oxygen was found to follow, for a considerable part, the crystal boundaries, so that presumably the cleavage strength σ_{cr} is reduced at the crystal boundaries. At first sight this seems to be peculiar, since a metallographic examination shows that the greater part of the oxygen in the metal is not along these boundaries but contained inside the crystals in the form of roughly spherical inclusions of iron oxide, and it is not to be supposed that these inclusions have any great effect upon the impact value. (It is known, for instance, that spherical inclusions of manganese sulphide have no appreciable effect upon impact strength, not even when they are present in large numbers.) Apparently it is to be assumed that some iron oxide is also present along the crystal boundaries, causing a considerable reduction of the intergranular cohesion. Remarkably enough, carbon has no great influence upon the impact value at 20 °C, even if this element is present at the crystal boundaries. This might be explained by assuming the oxide to be present as an almost continuous skin along the crystal boundaries and thus interrupting the metallic continuity over great distances, whilst the carbide may perhaps form isolated islands at the crystal boundaries. There are indications that something similar is

²⁾ J. D. Fast, Philips Techn. Rev. 11, 241-244, 1950 (No. 8).

the case in molybdenum³), which has the same crystal structure as iron.

Still more remarkable is the fact that mere traces of carbon (0.002 wt % or more) already reduce very considerably the harmful effect of oxygen. This is demonstrated by the upper curve in fig. 3, which shows, inter alia, that when the metal contains only a few thousandths per cent of carbon even an amount of 0.13% oxygen only reduces the impact value of iron from 17 to 13 kgm/cm². It is not yet clear how this action of traces of carbon is to be explained.

Impact value as a function of temperature

For determining the impact value of iron as a function of temperature an iron was used that had not such a high degree of purity as that of the "pure" iron referred to above; it was found to take too long to remove completely the last 0.01% of oxygen, with the aid of hydrogen, from the rather large quantities of iron (about 2 kg) needed for such measurements. The oxygen remaining after a comparatively short reduction time was therefore chemically bound by adding to the liquid metal the calculated quantity of pure zirconium, pure titanium or pure aluminium (see the article quoted in footnote²)).

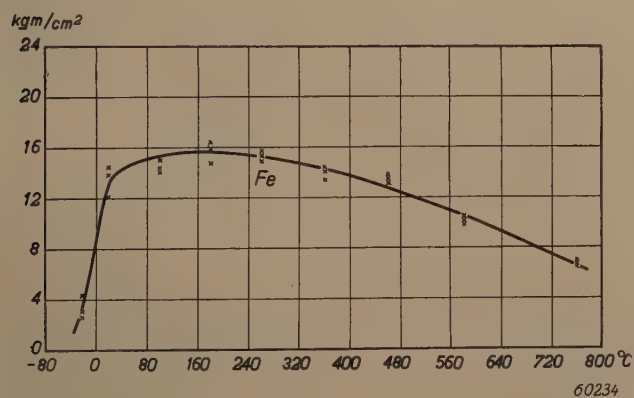


Fig. 4. Impact value of "pure" iron as a function of the temperature. The 0.012% oxygen contained in the iron has been bound with the equivalent quantity of zirconium to form zirconium oxide. There is a sharp rise in the impact-value curve at about 0 °C.

As an example fig. 4 gives the impact value, as a function of temperature, of iron in which the remaining 0.012% oxygen (all other impurities having been removed) was converted into zirconium oxide with the aid of 0.035% zirconium. The zirconium oxide remained in the metal in a finely divided state.

As is seen from fig. 4, the impact-value curve shows a comparatively sharp transition from high

to low values on the side of the low temperatures. For technical mild steel this sharp drop is well known. On the other hand the curve does not show the minimum that is known to occur with mild steel between 400 °C and 500 °C.

The drop round about 0 °C can be explained by the almost generally accepted assumption that as the temperature falls τ_{cr} increases at a relatively higher rate than σ_{cr} . As already mentioned, plastic deformation will take place so long as the relation $\sigma_{cr}/\tau_{cr} > \sigma_{max}/\tau_{max}$ is satisfied. According to the experiments (fig. 4) this is certainly the case for iron at high temperatures, but according to the above-mentioned assumption the ratio σ_{cr}/τ_{cr} decreases with falling temperature and ultimately becomes less than σ_{max}/τ_{max} . Roughly speaking, brittle fracturing will occur below the temperature at which the two relations become equal. The temperatures at which this will be the case will be all the higher as σ_{max}/τ_{max} is greater. Now from our previous discussions it appeared that this relation is much greater in the impact test than in the tensile test. Consequently the transition from deformation fracturing to brittle fracturing is to be expected at higher temperatures in the case of the impact test than in that of the tensile test. Such is indeed in agreement with what is found experimentally.

The quicker execution of the impact test compared with the tensile test is also to a small extent responsible for the fact that the transition range is shifted to higher temperatures, since the internal friction (τ_{cr}) increases with the rate of flow, so that under the conditions of the impact test one has to deal with smaller values of σ_{cr}/τ_{cr} than in the case of the tensile test.

The change from high to low impact values does not take place at any sharply defined transitional temperature but in a more or less wide range of temperatures. This can be explained by the fact that τ_{cr} increases with the degree of deformation, at least if it is permitted to assume (on which point there is no general agreement) that σ_{cr} changes less with deformation. Then, the ratio σ_{cr}/τ_{cr} decreases not only with falling temperature but also with increasing deformation. Consequently, at temperatures that are not much higher than the temperature T_k at which σ_{cr}/τ_{cr} and σ_{max}/τ_{max} in the non-deformed metal are equal, this equality will be reached and fracturing will take place after comparatively small deformations. According to these arguments, with rising temperature the impact value will change not abruptly but gradually from low to high values.

³) C. A. Zapffe, F. K. Landgraf and C. O. Worden, Metals Techn. 15, T. P. 2421, Aug. 1948.

dealt with, to an active element like zirconium, titanium or aluminium, then this element has a much more detrimental influence upon the mechanical properties of the metal. This is to be seen when comparing *fig. 6*, for iron containing 0.015% oxygen

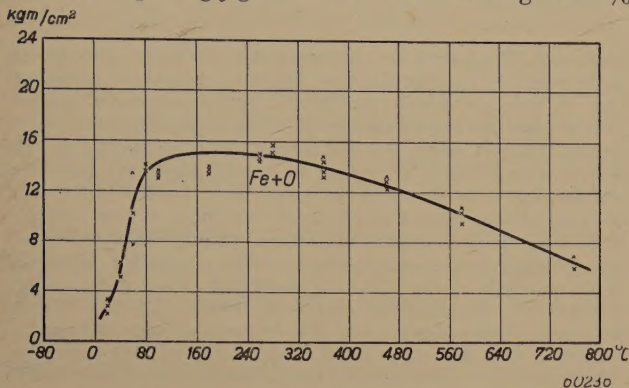


Fig. 6. Impact value of iron with 0.015% oxygen as the only impurity, as a function of temperature. (Here the oxygen is not bound as zirconium oxide.) The steep transition in the curve now lies at about 50 °C.

oxygen as the only impurity, with *fig. 4* relating to iron containing 0.012% oxygen in the form of ZrO_2 . The curve in *fig. 6* follows the same trend as that of the curve in *fig. 4* except that the transition range is shifted to higher temperatures. This agrees with the finding that oxygen present in the form of iron oxide reduces the σ_{cr}/τ_{cr} ratio of iron. Consequently as the temperature falls this ratio drops below the critical value at a higher value of the temperature. In view of the foregoing we presume the favourable effect of zirconium, aluminium, etc. to be due to these elements forming oxides which, contrary to the case with iron oxide, do not form continuous skins along the crystal boundaries.

Carbon

Fig. 7 relates to an iron containing 0.020%

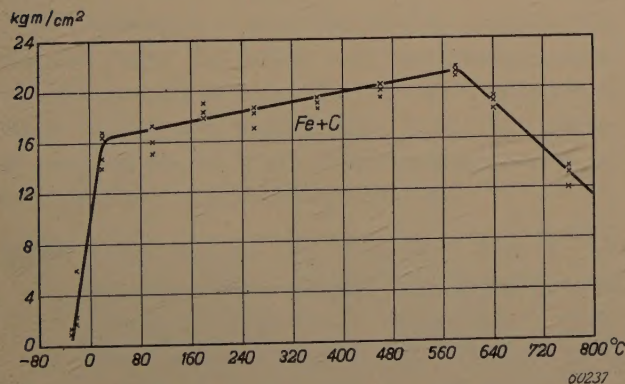


Fig. 7. Impact value of iron with 0.02% carbon as the only impurity, as a function of temperature. The steep transition lies at the same temperature as in *fig. 4* but the impact value continues to rise, though slowly, at higher temperatures. Above 580 °C the curve descends again as a result of the rapid recrystallization taking place during the deformation in the impact test.

carbon as the only impurity. (Here the oxygen is fairly easy to dispel, since when melting in vacuum it disappears in the form of CO .) Practically the same curve was found for iron with a carbon content four times as high. Contrary to what was found for "pure" and oxygenous iron, here the impact value between 20 °C and 580 °C still increases. Owing to more and more carbon passing into solution (solubility practically nil at 20 °C and equal to about 0.01% at 600 °C) apparently the shear strength decreases less quickly with rising temperature than it does in "pure" iron, so that the influence of the increasing deformed volume upon the amount of work taken up becomes predominant. Ultimately, above 580 °C the effect of recrystallization gains the upper hand.

Nitrogen

Nitrogen is but very slowly absorbed by "pure" iron, whereas it appears to be more quickly absorbed by oxygenous liquid iron and still more quickly

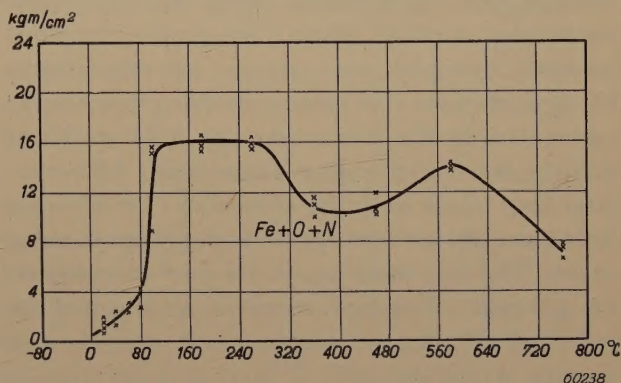


Fig. 8. Impact value, as a function of temperature, of iron containing 0.015% oxygen and 0.012% nitrogen as impurities.

by carbonaceous liquid iron (see the article quoted in footnote 2)). The influence of nitrogen upon the temperature-dependency of the impact value of iron has therefore not been separately investigated but deduced from the behaviour of iron containing O + N and that of iron containing C + N.

Fig. 8 gives the impact value, as function of temperature, of iron containing 0.015% oxygen and 0.012% nitrogen, whilst *fig. 9* relates to an iron containing 0.06% carbon and 0.018% nitrogen. Both curves show a flat minimum between 360 °C and 460 °C, which is apparently to be ascribed to the presence of nitrogen, since oxygen and carbon, each separately, do not give rise to any such a minimum (cf. *figs 6* and *7*).

Although carbon and nitrogen, when present in steel as impurities, generally behave as equivalent elements, they appear to show a striking difference

here. No entirely satisfactory explanation can yet be given for the occurrence of the minimum in the impact-value curve due to the presence of nitrogen.

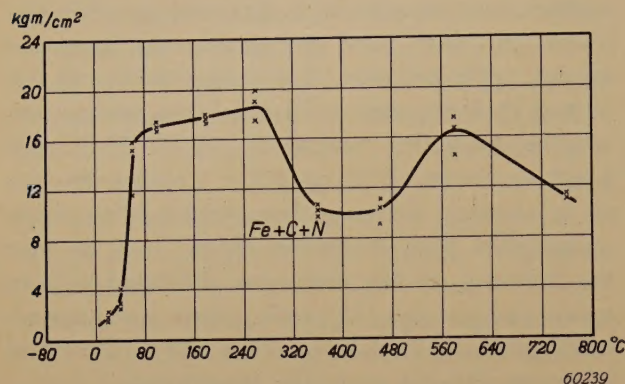


Fig. 9. Impact value, as a function of temperature, of iron containing 0.06% carbon and 0.018% nitrogen as impurities. As in the previous curve, a minimum occurs between 360 $^{\circ}\text{C}$ and 460 $^{\circ}\text{C}$, which is to be ascribed to the presence of nitrogen.

Furthermore, experiments carried out with notched bars of technical mild steel (i.e. iron containing, among others, manganese, sulphur, phosphor, carbon and oxygen as admixtures) likewise showed a minimum in the impact-value curve, also in the case where the mild steel was entirely freed of nitrogen. Experiments with mild steel from which both the carbon and the nitrogen had been driven out gave a flat impact-value curve. Whereas, therefore, in the case of pure iron the presence of nitrogen seems to be essential for

the occurrence of a minimum, in the case of mild steel apparently carbon is able to take over the part played by nitrogen.

Summary. The results are discussed of experiments carried out to investigate the notched-bar impact strength of pure iron and of iron to which impurities were purposely added. This is preceded by a description of the impact test according to the Charpy method and a review of the simplest concepts that can be formed according to known theories about the relation between the state of stress in a loaded metal and the possible occurrence of plastic deformation and/or rupture. According to this simplified picture, plastic deformation of the material takes place when under increasing load the shear strength of the metal is exceeded before the cleavage strength; in the opposite case the metal shows a brittle fracture. In the impact test the state of stress during the loading is such that there is a greater chance of brittle fracture than under the conditions of the ordinary tensile test. — It appears that oxygen reduces the impact value of iron at 20 $^{\circ}\text{C}$ considerably. Traces of carbon may greatly diminish this harmful effect of oxygen. As a result of the marked increase of shear strength with falling temperature (assuming the cleavage strength to remain constant) the impact value of pure iron when the temperature drops to around 0 $^{\circ}\text{C}$ shows the same transition from high to low values as is already known in the case of normal technical mild steel. Due to an adiabatic elevation of temperature taking place in the deformation zone during the bending of the test bar, this transition extends over a shorter temperature range than is the case with tensile or bending tests carried out more slowly. When oxygen is present in the iron the transition range is shifted to higher temperatures: oxygen reduces the cleavage strength of iron. Above the transition range carbon causes the impact-value curve to rise gradually with increasing temperature, whilst nitrogen gives rise to a flat minimum in the curve at temperatures of about 400 $^{\circ}\text{C}$. Such a minimum is shown by the impact-value curve of mild steel, also when it does not contain any nitrogen. No really satisfactory explanation for this phenomenon can yet be given.

ABSTRACTS OF RECENT SCIENTIFIC PUBLICATIONS OF THE N.V. PHILIPS' GLOEILAMPENFABRIEKEN

Reprints of these papers not marked with an asterisk can be obtained free of charge upon application to the Administration of the Research Laboratory, Kastanjelaan, Eindhoven, Netherlands.

1870*: G. Diemer and K. S. Knol: Measurements on total-emission conductance at 35 cm and 15 cm wavelengths (*Physica* **15**, 459-462, 1949, No. 5/6).

From Q-measurements on a circuit containing a disc-seal diode, the electronic conductance was calculated and plotted as a function of $-V_a$. The experimental results do not at all agree with the linear field theory. The influence of space charge is discussed. In the space-charge-limited region the total emission conductance does not play such an important part as might be expected from an extrapolation of the g -values measured beyond cut-off.

1871*: K. S. Knol and A. Versnel: Suppression of shot effect noise in triodes and pentodes (*Physica* **15**, 462-464, 1949, No. 5/6).

It is shown experimentally that the presence of a capacitance between the control grid and the anode results in a much larger suppression of shot effect noise by detuning, as suggested by Strutt and Van der Ziel, than is possible in the absence of such a capacitance, as may be expected from a theoretical point of view.

1872: J. de Jonge and R. Dijkstra: The photo-decomposition products of some diazonium salts (*Rec. Trav. chim. Pays-Bas* **68**, 424-429, 1949, No. 6).

By irradiation of solutions of some benzene diazonium salts (including one *p*-hydroxybenzene diazonium salt) in water the corresponding phenols could be isolated with a high yield. The photo-decomposition of two *p*-diazo-oxides seems to have a more complicated character.

1873: J. de Jonge, R. Dijkstra and P. B. Braun: The thermal decomposition of *o*-hydroxy-diazonium compounds (*Rec. Trav. chim. Pays-Bas* **68**, 431-432, 1949, No. 6).

Experimental evidence has been found that the thermal decomposition of *o*-hydroxy-diazonium salts may be identical with the photo-decomposition.

1874/75: B. H. Schultz: On the application of Reynolds' analogy and the heat-exchange factor to the design of heat exchangers, I-III (*Appl. sci. Res.* **A1**, 387-416, 1949, No. 5/6).

Part. I. Reynolds's analogy between heat transfer and friction leads to the well-known equation $h = \frac{1}{2} f C_p \rho v$. According to improved theories, for turbulent flow of gases in long straight ducts, the factor $\frac{1}{2}$ is to be replaced by $c = 0.55$. It is shown that for linear flow in long ducts c is equal to 0.3 for a circular and 0.4 for a flat rectangular cross-section. Further, in very short ducts an approximate theory is shown to give values that are of the same order of magnitude. Though the equation $h = c f C_p \rho v$ has an approximative character, it has the advantage of being simple and generally valid for gases flowing in ducts.

Part II. For heat exchangers in which the temperature differences between gas and wall decrease considerably it proves to be useful to introduce a "heat-exchange factor" K . A simple relation between K and the pressure drop can be derived from the modified form of Reynolds's analogy given in part I. This relation is applied to some problems of optimum dimensions for heat exchangers.

Part III. The value of c in the relation between K and the pressure drop for turbulent flow is taken from the experimental data given in the literature. The author's measurements in the transition region show that, though both heat exchange and friction may vary considerably with changing flow conditions at (or before) the entrance of the ducts, the factor c does not vary appreciably.

In those cases where the curve for f shows a pronounced "dip" this is also found in the curve for the heat exchange factor.

1876: J. A. Haringx: Elastic stability of helical springs at a compression greater than the original length (*Appl. sci. Res.* **A1**, 418-434, 1949, No. 5/6).

A former statement that, as to the behaviour of helical compression springs in respect of elastic stability, it is quite immaterial whether the ends

of the springs are hinged, constrained parallel or clamped, appears to be true only for compressions less than the original spring length. For greater compressions, on the other hand, the differences occurring for the various end conditions are found to be quite characteristic. Since a (cylindrical) spring could be realized with a compression greater than the original length, simply by turning it inside out, the new theoretical results found for this region of compressions could be verified by experiment. These experiments demonstrated the remarkable toppling over of the coils which accompanies the transition from the stable to the unstable state. Moreover they confirmed the existence of the characteristic differences in the behaviour of the spring for the various end conditions, though qualitatively only. An explanation for the numerical deviations occurring is mentioned.

1877: A. Bremmer: The propagation of electromagnetic waves through a stratified medium and its W.K.B. approximation for oblique incidence (*Physica* 15, 593-608, 1949, No. 7).

The plane-wave solution of Maxwell's equations for a stratified medium is split into a series of terms with a simple geometrical meaning, the first of which constitutes the W.K.B. approximation. The original vector problem is reduced to a scalar problem by introducing a convenient hertzian vector. The application of the saddle-point method to the individual terms of the series mentioned leads to simple geometric-optical approximations.

1878: F. A. Kröger: Sodium and lithium as activators of fluorescence in zinc sulphide (*J. Opt. Soc. Amer.* 39, 670-672, 1949, No. 8).

Lithium and sodium dissolved in zinc sulphide act either as activators of fluorescence or as quenchers, according to whether chlorine is present or not. The fluorescence occurs only at low temperatures and is excited by short-wave ultraviolet. The lithium band has a maximum at 4380 Å in the sphalerite modification of ZnS; the sodium band has a maximum at 3940 Å in the sphalerite and at 3800 Å in the wurtzite modification.

The difference in the position of the bands for wurtzite and sphalerite and a shift found upon incorporation of CdS proves that we are dealing with electron-transfer bands. The centers of fluorescence are supposed to consist of lithium or sodium and chlorine ions occupying normal lattice sites. The levels involved in the fluorescence transitions are levels due to lattice ions, changed by the presence of the monovalent ions (indirect activation).

The quencher centers are supposed to consist of Li_2S and Na_2S incorporated in ZnS in such a way that the alkaline ions occupy two cation sites, while of two anion sites one is occupied by a monovalent sulfur ion and the other by an electron.

1879: G. W. Rathenau and H. de Wit: Indicator for small amounts of oxygen in reducing furnace atmospheres (*Metallurgia* 40, 114, 1949, June).

When annealing metals or alloys in pure hydrogen a rough determination of the oxygen pressure of the protecting gas is often desirable. This is done by means of an electrically heated Cr-Ni-steel strip, which shows surface oxidation below the temperature at which the oxygen pressure of the atmosphere equals that of the chromium oxide.
